The Reliability of an Aerobic and an Anaerobic Exercise Tolerance Test in Patients with Juvenile Onset Dermatomyositis

TIM TAKKEN, JANJAAP van der NET, and PAUL J.M. HELDERS

ABSTRACT. Objective. To investigate the reliability of an aerobic and an anaerobic exercise test in patients with juvenile dermatomyositis (JDM).

Methods. Sixteen patients with JDM (mean age 13.85 ± 6.4 yrs, range 6.7–27.2) participated. Anaerobic exercise capacity was measured using the Wingate Anaerobic Exercise Test (WAnT). Aerobic exercise test was performed using a graded exercise test to volitional exhaustion on an electronically braked cycle ergometer. Patients were tested and retested within 12.8 ± 5.7 days.

Results. Correlation coefficients and other reliability statistics indicated that peak power and mean power on the WAnT and peak oxygen uptake (VO$_{2\text{peak}}$) and maximal workload (W$_{\text{max}}$) on the aerobic exercise test could be reliably assessed in patients with JDM. Pearson (R) and intraclass correlation coefficients (ICC) were > 0.85, and typical error was < 20% for the peak power and mean power on the WAnT. On the aerobic exercise test the R and ICC were > 0.95, and typical error was < 6% for the VO$_{2\text{peak}}$ and W$_{\text{max}}$.

Conclusion. We found acceptable reliability for the WAnT and very good reliability for the aerobic exercise test. This indicated that these exercise tests could be performed reliably in patients with JDM. (J Rheumatol 2005;32:734–9)

Key Indexing Terms:
SHORT-TERM MUSCLE POWER  EXERCISE TOLERANCE  EXERCISE TEST METHODS
CHILDREN  ADOLESCENTS  JUVENILE DERMATOMYOSITIS

Traditionally, muscle strength was the principal component of interventions in inflammatory myositis$^{1,2}$. Recently, clinicians have focused on exercise capacity as well$^{3-6}$. These and other studies revealed disturbances in muscle metabolism in patients$^{7-10}$.

Clinical exercise physiology is a rapidly growing discipline within healthcare. As a consequence, more physiological instruments originally designed for use in a healthy population are being applied in a clinical population. In patients with juvenile dermatomyositis (JDM), aerobic exercise tests have been used in the management and evaluation of health status$^{3,4}$. These exercise tests were able to determine the aerobic exercise capacity (i.e., peak oxygen uptake, VO$_{2\text{peak}}$). Besides aerobic physical fitness, recent reports suggest the importance of anaerobic physical fitness (intensive exercise lasting < 2 minutes). Since many daily childhood activities consist of short-term bursts of intensive activity, anaerobic fitness is an important measure$^{11}$. In patients with juvenile idiopathic arthritis (JIA) it has been shown that anaerobic physical fitness was better correlated with functional ability compared to aerobic fitness, suggesting that anaerobic exercise capacity might be more important for functional ability in patients with a pediatric rheumatic disease$^{12}$.

The Wingate Anaerobic Exercise Test (WAnT) has been used in healthy children$^{13}$, in patients with neuromuscular diseases$^{14}$, and in those with JIA$^{12,15}$, and found to be reliable and valid compared to other anaerobic indices such as those invasively determined (by muscle biopsy), i.e., muscle metabolites, energy-rich phosphates, and muscle fiber composition$^{16}$. In a previous study we described the feasibility of the WAnT in patients with JDM and juvenile polymyositis and found a decrease in their short-term anaerobic exercise capacity of approximately 30% compared to healthy controls$^{17}$. In another study we found a 34% decreased aerobic exercise capacity during treadmill exercise testing in patients with JDM$^{3}$. However, the reliability of both the aerobic and anaerobic exercise tests has never been determined in patients with JDM. Since reliability is an important issue for clinical use (followup as well as clinical trials), we investigated the reliability of an aerobic and an anaerobic exercise test in patients with JDM.

From the Department of Pediatric Physical Therapy and Exercise Physiology, University Hospital for Children and Youth, Het Wilhelmina Kinderziekenhuis, University Medical Center Utrecht, Utrecht, The Netherlands.

Supported by the Dutch Arthritis Association (NR 02-2-102).

T. Takken, MSc, PhD; J. van der Net, PT, PhD; P.J.M. Helders, PT, MSc, PhD.

Address reprint requests to Dr. T. Takken, Department of Pediatric Physical Therapy and Exercise Physiology, University Hospital for Children and Youth, Het Wilhelmina Kinderziekenhuis, University Medical Center Utrecht, Room KB.02.056, PO Box 85090, 3508 AB Utrecht, The Netherlands. E-mail: t.takken@wkz.azu.nl

Accepted for publication December 20, 2004.
**MATERIALS AND METHODS**

**Patients.** Sixteen patients with JDM participated in this study. Patients were recruited from the pediatric rheumatology outpatient clinic of the Wilhelmina Children’s Hospital of the University Medical Center Utrecht and fulfilled the criteria for the diagnosis of JDM as described by Bohan and Peter18. The patients were considered to be in a stable clinical condition. Their characteristics are described in Table 1. Patients and parents gave their informed consent for participating in the study. The Institutional Review Board approved all procedures. Patients were tested and retested within 12.8 ± 5.7 days during the summer of 2003. All tests were performed by the same clinical exercise physiologist (TT), with 6 years’ experience in exercise testing of patients with a pediatric rheumatic disease.

**Anthropometry.** The participants’ body mass and height were determined using an electronic scale and a measuring stick. Body composition was assessed using the sum of 7 skin folds (ΣSF) method according to Pollack, et al19. The measurements were taken at 7 sites (on the right side of the body: triceps, biceps, subscapular, suprailliac, mid-abdominal, medial calf, and thigh) by the investigator (TT) in accord with the American College of Sports Medicine guidelines20.

**Anaerobic exercise capacity.** The WAnT was performed as described by Bar-Or13 on a recently calibrated electromagnetic braked cycle ergometer (Lode Examiner, Lode BV, Groningen, The Netherlands). The ergometer was upgraded and calibrated by the manufacturer to a maximal resistance of 800 W instead of the standard 400 W. The external resistance was controlled and the power output was measured using the Lode Wingate software package21. The seat height was adjusted to the patient’s leg length (comfortable cycling height).

The external load (torque; in N•m) was determined, dependent on body weight (at 0.53 × body weight and 0.55 × body weight for girls and boys under 14 years of age, and 0.67 × body weight and 0.7 × body weight for older girls and boys, respectively) according to the user manual.

The patients were instructed to exercise for 1 min at the cycle ergometer with an external load of 15 W at 50 RPM. Then the sprint protocol started; patients were instructed to cycle all-out for 30 s. The measured variables were mean power, peak power, lowest power, and fatigue index. Mean power represented the average power output over the 30 s sprint. Peak power was the highest recorded power output achieved during the 30 s sprint, lowest power was the lowest power recorded during the 30 s sprint. The fatigue index represented the decline in power output (fatigue) during the 30 s sprint, and was calculated as highest power minus lowest power divided by the time interval between highest power and lowest power. Power output during the WAnT was corrected for the inertia of the mass of the flywheel (23.11 kg•m²).

**Aerobic exercise test.** Subjects performed an aerobic exercise test using an electronically braked cycle ergometer (Lode Examiner). Three of the youngest patients did not fit this ergometer; they were tested on a smaller mechanically braked ergometer (Tunturi, Tampere, Finland).

The seat height of the ergometer was adjusted to the patient’s leg length. One minute of unloaded cycling preceded the application of resistance to the ergometer. Then the workload was increased by constant increment of 20 W every minute. This protocol continued until the patient stopped because of volitional exhaustion, despite strong verbal encouragement from the investigators. The highest achieved workload during the test (Wmax) was recorded.

During the aerobic exercise test, subjects breathed through a facemask (Hans Rudolph Inc., Kansas City, MO, USA) connected to a calibrated metabolic cart (Oxycon Champion, Jaeger, Viasys, Bilthoven, The Netherlands). Expired gas was passed through a flow meter, an oxygen (O2) analyzer, and a carbon dioxide (CO₂) analyzer. The flow meter and gas analyzers were connected to a computer, which calculated breath-by-breath minute ventilation (Ve), oxygen consumption (VO₂), carbon dioxide production (VCO₂), and respiratory exchange ratio (RER) from conventional equations. Heart rate (HR) was measured continuously during the aerobic exercise test by a bipolar electrocardiogram. Peak oxygen consumption (VO₂peak) was taken as the average value over the last 30 s during the aerobic exercise test. Heart rate recovery (HRrecov) was calculated as the maximum HR (HRmax) attained during the exercise test minus the HR after 1 min recovery.

**Statistics.** The data were analyzed using SPSS 11.0 and MS Excel 98 for Windows. Correlation coefficients [Pearson correlations (R) and intraclass correlations (ICC)] were computed for test-retest reliability. Acceptable reliability criteria for ICC values were values > 0.7522.

Moreover, limits of agreement (LOA) were calculated to conform to the procedure described by Bland and Altman23. Bland-Altman analysis described the level of agreement between 2 measurements23. In this analysis, the “bias” is an estimate of how closely on average the 2 measurements agree and the “precision” indicates how well the methods agree for an individual23. The LOA are calculated by multiplying the precision by 1.9623. Typical error and total error were calculated according to Hopkins24. Typical error was calculated as the standard deviation in each subject’s measurements between tests, after any shifts in the mean have been taken into account. Here the typical error is expressed as a percentage of the subject’s mean score to obtain an interpretable percentage score. This percentage is also known as the coefficient of variation24. Total error was calculated as the mean of each subject’s standard deviation between the trials25. The level of statistical significance was set at p = 0.05.

**RESULTS**

Three of the youngest patients did not fit on the electromagnetic braked ergometer, therefore only data of 13 patients were available for analysis of the reliability of the WAnT. However, the youngest patients were able to perform the aerobic exercise test when they were tested on a suitable size mechanically braked ergometer (see Methods).

The physiological variables measured during the exercise test on both occasions are described in Table 2. The reliability statistics of both the WAnT and the aerobic exercise test are given in Table 3. As shown by Bland-Altman plots (Figures 1 and 2), there was one obvious outlier. Even when this outlier was included in the calculations, the reliability statistics are still acceptable. We found R and ICC values for mean power, peak power, and fatigue index were ± 0.85 or above. The typical error (expressed as a percentage) showed that peak power and mean power were the 2 variables of the WAnT with the lowest error of measurement. The lowest power and the fatigue index had almost twice as much variability compared to mean and peak power.

When this outlier was omitted from the analysis, the reliability statistics for peak power were R = 0.938, ICC = 0.937, LOA = 137.86 W, typical error = 49.77 W, total error = 58.40 W, and typical error = 14.8%. The reliability statis-
tics for mean power after omission of this patient were \( R = 0.974, ICC = 0.972, LOA = 47.25 \text{ W}, \) typical error = 17.07 W, total error = 21.9 W, and typical error was 8.5%. The reliability statistics of the aerobic exercise test showed a moderately low variability (Figure 3, Figure 4, Table 3). \( \text{VO}_{2\text{peak}}, \text{Wmax} \) showed \( R \) and ICC > 0.95, which indicated a very good reliability. \( \text{HR}_{\text{max}}, \text{HR}_{\text{recov}} \) had \( R \) and ICC values between 0.59 and 0.78, indicating moderate reliability compared to the other measured variables. In the \( \text{VO}_{2\text{peak}} \) test, there was either no change (\( \text{VO}_{2\text{peak}} \)) or a very small change in the mean (\( \text{Wmax}, \text{RER}_{\text{max}}, \text{HR}_{\text{max}}, \text{HR}_{\text{recov}} \)), which indicated that there was no systematic learning effect of this test (Table 2).

**DISCUSSION**

This is the first study investigating the reliability of exercise tests in patients with JDM. We found that both anaerobic and aerobic exercise tests were reliable exercise tests in patients with JDM. Tirosh, et al\(^{14} \) found very good reliability (\( R > 0.95 \)) of the WAnT in a variety of patients with disorders such as Duchenne muscular dystrophy, Becker’s muscular dystrophy, cerebral palsy, and spinal muscular atrophy. When the one outlier was omitted from the analysis we found comparable results in patients with JDM. This patient was known to have a behavioral dysfunction, and he sometimes lacked the motivation to perform in clinical tests. This case highlights that exercise tests are influenced by the motivation of the patient to give a maximal effort and that lack of motivation can influence the final test result.

The change in mean in both mean power and peak power during the WAnT showed that both increased during the second assessment, while the lowest power decreased and consequently the fatigue index increased. This result might be because one of the patients scored significantly higher in the second assessment compared to the first. When this patient was omitted from the analysis, the change in mean was reduced to 18 W for peak power and 10 W for mean power. When using the WAnT as an outcome for an intervention study, one should consider a learning session to rule out this small systematic learning effect, which could affect the power of a study.

The reliability statistics of the aerobic exercise test showed that this test was very accurate in determining \( \text{VO}_{2\text{peak}} \) and \( \text{Wmax} \) in patients with JDM. In previous studies in patients with juvenile arthritis, error of measurement in \( \text{VO}_{2\text{peak}} \) of 10% to 12% were found.\(^{26,27} \) In our sample of patients with JDM the error of measurement was 5.5%. This difference might reflect the day-to-day variation in disease expression in patients with juvenile arthritis, while patients with JDM might have less short-term fluctuation in their disease state. The total error of \( \text{VO}_{2\text{peak}} \) was 0.8 l \( \times \) \( \text{min}^{-1} \), and was well within the quality range of certified exercise physiology laboratories, and indicates the reliability of the aerobic exercise test. Our error level was comparable with observations in healthy children.\(^{29} \)

Since JDM is a very rare disease, our sample size was small. To improve sample size, multicenter studies could be

---

**Table 2. Reproducibility of performance in the duplicate anaerobic and aerobic exercise tests.**

<table>
<thead>
<tr>
<th>Measurement 1</th>
<th>Measurement 2</th>
<th>Change in Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Peak power, W</td>
<td>442.23 (183.9)</td>
<td>481.92 (212.7)</td>
</tr>
<tr>
<td>Mean power, W</td>
<td>244.9 (93.4)</td>
<td>270.0 (113.8)</td>
</tr>
<tr>
<td>Lowest power, W</td>
<td>132.7 (62.8)</td>
<td>117.62 (55.7)</td>
</tr>
<tr>
<td>Fatigue index</td>
<td>12.45 (6.6)</td>
<td>14.8 (7.3)</td>
</tr>
<tr>
<td>( \text{VO}_{2\text{peak}}, \text{Wmax} )</td>
<td>1.3 (0.44)</td>
<td>1.3 (0.39)</td>
</tr>
<tr>
<td>( \text{Wmax} )</td>
<td>101.7 (38.8)</td>
<td>104.7 (41.9)</td>
</tr>
<tr>
<td>( \text{RER}_{\text{max}} )</td>
<td>1.14 (0.11)</td>
<td>1.14 (0.06)</td>
</tr>
<tr>
<td>( \text{HR}<em>{\text{max}}, \text{HR}</em>{\text{recov}}, \text{beats} \times \text{min}^{-1} )</td>
<td>178.4 (16.5)</td>
<td>175.3 (14.7)</td>
</tr>
<tr>
<td>( \text{HR}_{\text{recov}}, \text{beats} \times \text{min}^{-1} )</td>
<td>51.13 (22.67)</td>
<td>48.25 (18.72)</td>
</tr>
</tbody>
</table>

Change in mean denotes the change between measurement 1 and measurement 2. \( \text{VO}_{2\text{peak}}, \text{peak oxygen uptake}, \text{Wmax}: \text{maximal workload}, \text{RER}_{\text{max}}, \text{maximal respiratory exchange ratio}, \text{HR}_{\text{max}}: \text{maximal heart rate}, \text{HR}_{\text{recov}}, \text{recovery heart rate.} \)

---

**Table 3. Reliability statistics of the anaerobic and the aerobic exercise tests.**

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>ICC</th>
<th>Typical Error</th>
<th>Total Error</th>
<th>LOA</th>
<th>Typical Error, % (CV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak power, W</td>
<td>0.879</td>
<td>0.869</td>
<td>71.88</td>
<td>91.38</td>
<td>199.11</td>
<td>18.7</td>
</tr>
<tr>
<td>Mean power, W</td>
<td>0.860</td>
<td>0.851</td>
<td>41.17</td>
<td>54.76</td>
<td>114</td>
<td>16.8</td>
</tr>
<tr>
<td>Lowest power, W</td>
<td>0.735</td>
<td>0.73</td>
<td>30.84</td>
<td>31.06</td>
<td>85.42</td>
<td>41.0</td>
</tr>
<tr>
<td>Fatigue index</td>
<td>0.849</td>
<td>0.845</td>
<td>2.74</td>
<td>3.79</td>
<td>7.6</td>
<td>28.3</td>
</tr>
<tr>
<td>( \text{VO}<em>{2\text{peak}}, \text{Wmax}, \text{RER}</em>{\text{max}}, \text{HR}<em>{\text{max}}, \text{HR}</em>{\text{recov}}, \text{beats} \times \text{min}^{-1} )</td>
<td>0.962</td>
<td>0.955</td>
<td>0.09</td>
<td>0.08</td>
<td>0.24</td>
<td>5.8</td>
</tr>
<tr>
<td>( \text{Wmax}, \text{W} )</td>
<td>0.969</td>
<td>0.969</td>
<td>7.28</td>
<td>4.0</td>
<td>20.15</td>
<td>3.6</td>
</tr>
<tr>
<td>( \text{RER}_{\text{max}} )</td>
<td>0.677</td>
<td>0.589</td>
<td>0.06</td>
<td>0.05</td>
<td>0.15</td>
<td>4.8</td>
</tr>
<tr>
<td>( \text{HR}_{\text{max}}, \text{beats} \times \text{min}^{-1} )</td>
<td>0.904</td>
<td>0.898</td>
<td>4.99</td>
<td>4.27</td>
<td>13.82</td>
<td>2.4</td>
</tr>
<tr>
<td>( \text{HR}_{\text{recov}}, \text{beats} \times \text{min}^{-1} )</td>
<td>0.78</td>
<td>0.767</td>
<td>10.04</td>
<td>10.61</td>
<td>27.8</td>
<td>28.8</td>
</tr>
</tbody>
</table>

\( R: \text{Pearson product moment correlation}, \text{ICC: intraclass correlation coefficient}, \text{LOA: limits of agreement}, \text{CV: coefficient of variation}, \text{VO}_{2\text{peak}}, \text{peak oxygen uptake}, \text{Wmax}: \text{maximal workload}, \text{RER}_{\text{max}}, \text{maximal respiratory exchange ratio}, \text{HR}_{\text{max}}: \text{maximal heart rate}, \text{HR}_{\text{recov}}, \text{recovery heart rate.} \)
initiated. However, this could introduce bias due to the use of different equipment.

The introduction of clinical exercise physiology as a diagnostic and therapeutic tool in the management of JDM seems promising. Not only was aerobic and anaerobic testing found to be reliable, in our clinical practice it also is feasible to conduct both exercise testing modes in patients with JDM over a large age span. This will facilitate followup of these patients well into adulthood, since most commonly used outcome measures (e.g., the Childhood Health Assessment Questionnaire and Childhood Myositis Assessment Score) are age-limited, as they are designed and validated for use in children and adolescents only. However, exercise tests have been successfully performed over a very wide age span, from as young as 5 years of age\textsuperscript{30} to 75 years\textsuperscript{31}, and thus will facilitate the longitudinal followup of our patients. Future studies should determine the sensitivity to change of both anaerobic and aerobic exercise tests during exercise therapy and medical treatment in patients with JDM.

We found acceptable reliability for the WAnT and very good reliability for the aerobic exercise test ($\text{VO}_2\text{peak}$ and $W_{\text{max}}$). This indicates that exercise tests can be reliably performed in patients with JDM.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Bland–Altman plot of peak power during test and retest on the WAnT. Bold line shows the mean difference between the 2 measurement methods, the 2 thin lines indicate $\pm$ 2 SD. X axis: average peak power value from both tests. Y axis: difference between peak power during the test and peak power during the retest.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Bland–Altman plot of mean power during test and retest on the WAnT. Bold line shows the mean difference between the 2 measurement methods, the 2 thin lines indicate $\pm$ 2 SD. X-axis: average mean power value from both tests. Y axis: difference between mean power during the test and mean power during the retest.}
\end{figure}
ACKNOWLEDGMENT

We thank Susan Maillard, MSc, for her review of the manuscript.

REFERENCES


Figure 3. Bland–Altman plot of VO_{2peak} during test and retest. Bold line shows mean difference between the 2 measurement methods, the 2 thin lines indicate ± 2 SD. X axis: average VO_{2peak} value from both tests. Y axis: difference of VO_{2peak} during the test and the retest.

Figure 4. Bland–Altman plot of W_{max} during test and retest. Bold line shows mean difference between the 2 measurement methods, the 2 thin lines indicate ± 2 SD. X axis: average W_{max} value from both tests. Y axis: difference of W_{max} during the test and the retest.


