Reproducibility of Cardiopulmonary Exercise Measurements in Patients With Pulmonary Arterial Hypertension*

James E. Hansen, MD; Xing-Guo Sun, MD; Yuji Yasunobu, MD, PhD; Robert P. Garafano, PhD; Gregory Gates, EdM; Robyn J. Barst, MD; and Karlman Wasserman, PhD, MD

Background and objectives: As part of a recent study, cardiopulmonary exercise tests (CPETs) were used to evaluate and follow up patients with pulmonary arterial hypertension (PAH). These patients were more impaired than those in other published series evaluating CPET reproducibility. We used these patient tests to assess patient performance variability and evaluate reading variability. To achieve this end, six independent evaluators graded key CPET measurements in patients with PAH who underwent duplicate CPETs within 3 days of each other.

Setting and patients: Over a 15-month period at two tertiary-care teaching hospitals, 42 patients with PAH underwent repeated, paired CPETs using cycle ergometry.

Interventions and measurements: Each patient underwent one to six pairs of cycle ergometry tests to maximal tolerance. Each pair of tests was separated by 3 months, with each test in the pair separated by 1 to 3 days. Specific guidelines were given to the independent evaluators for the key measurements assessed from each CPET study: peak O2 uptake (V˙O2), peak heart rate, peak O2 pulse, anaerobic threshold (AT), and end-tidal PO2, end-tidal PCO2, and the ventilatory equivalent for CO2 at the AT (V˙E/V˙CO2@AT).

Results: There were no fatalities or complications occurring among the 242 tests performed on 42 patients. The mean peak V˙O2 was 722 mL/min or 41% of predicted; 34 patients were Weber class C or D. Using the specific guidelines to measure the variability of measurements made by the six independent evaluators, the coefficients of variation were <2.2% for peak V˙O2, peak heart rate, peak O2 pulse, anaerobic threshold values at the AT, and V˙E/V˙CO2@AT, while for the AT, it was 8.5%. There were no significant differences in these measurements between the first and second tests of any pair or between the earlier and later sets of pairs.

Conclusions: Using specific guidelines, key CPET measurements can be safely, reliably, and reproducibly assessed even in patients with severe exercise intolerance.

(CHEST 2004; 126:816–824)

Key words: anaerobic threshold; exercise testing; O2 pulse; peak O2 uptake; reproducibility; ventilatory efficiency; ventilatory equivalent for CO2

Abbreviations: AT = anaerobic threshold; CoV = coefficient of variation; CPET = cardiopulmonary exercise test; peak HR = heart rate at peak O2 uptake; PAH = pulmonary arterial hypertension; peak O2P = O2 pulse at peak O2 uptake; PCO2@AT = mixed-expired pressure of CO2 at and immediately after the anaerobic threshold; PETCO2@AT = end-tidal pressure of CO2 at and immediately after the anaerobic threshold; V˙CO2 = CO2 output; V˙E/V˙CO2@AT = ventilatory equivalent for CO2 at and immediately after the anaerobic threshold; V˙O2 = O2 uptake

In normal children1 and adults,2,3 in highly trained athletes,2,4,5 and in patients with moderate exercise impairment due to chronic disease,6–24 repeated measurements obtained during cardiopulmonary exercise tests (CPETs) have generally shown good reliability and reproducibility. Particularly important measurements are peak O2 uptake (peak V˙O2), which defines the maximal transport of extractable O2 to the tissues; heart rate at peak V˙O2 (peak HR); O2

*From the Department of Medicine (Drs. Hansen, Sun, Yasunobu, and Wasserman), Research and Education Institute, Harbor-UCLA Medical Center, Torrance, CA; and Department of Pediatrics (Drs. Garafano, Gates, and Barst), Columbia Presbyterian Medical Center, New York, NY.

Manuscript received October 7, 2003; revision accepted May 6, 2004.

Reproduction of this article is prohibited without written permission from the American College of Chest Physicians (e-mail: permissions@chestnet.org).

Correspondence to: James E. Hansen, MD, Box 405, Department of Medicine, Harbor-UCLA Medical Center, Torrance, CA 90509; e-mail: jhansen@rei.edu
pulse at peak $\dot{V}O_2$ (peak $O_2$P), which reflects peak stroke volume and cardiac output; anaerobic threshold (AT), which assesses maximal sustainable aerobic work capacity; ventilatory equivalent for $CO_2$ at the AT ($Ve/\dot{V}CO_2$@AT), which is a good estimate of gas exchange abnormalities and is often described as ventilatory efficiency; and end-tidal $CO_2$ value near the AT ($Pe\text{t}CO_2$@AT) and end-tidal $O_2$ value near the AT ($Pe\text{t}O_2$@AT), which also assess gas exchange and alveolar ventilation.

Severe disability could increase the variability and decrease the confidence of interpretation of these key CPET measurements, as suggested in one series that showed higher measurement variability in comparing severe to less severe heart failure. The paired CPET data from a number of patients with pulmonary artery hypertension (PAH) evaluated as part of a therapeutic study at two sites were available for evaluation. Because of the overall severity of these patients’ illness, whether assessed by Weber classification or percentage of predicted peak $\dot{V}O_2$, we thought it important to evaluate the reproducibility of their key CPET measurements to ascertain if their single test values were as reliable as CPET values from less disabled patients in other reported series. Additionally, in this population, we recognized that right-to-left shunting of $CO_2$-rich blood through a patent foramen ovale, which frequently occurs at the onset of exercise in patients with PAH, might interfere with the assessment of the AT.

Because commercial breath-by-breath systems have differing time frames for selecting these key variables, we established predetermined rules for the use of the six study-blinded independent evaluators to record their measurements. Using paired CPETs conducted 1 to 3 days apart, we assessed not only the variability between these measurements in 228 tests conducted 1 to 3 days apart, we assessed not only the variability between these measurements in 228 tests (114 paired) from 42 patients (observer plus patient performance variability), but also the variability of evaluators in measuring key measurements in a single test (interobserver variability).

**Materials and Methods**

**Subjects**

This study was carried out at two sites: Harbor-UCLA Medical Center, Torrance, CA, and Columbia Presbyterian Medical Center, New York, NY. Patients with PAH had been consecutively enrolled in a therapeutic drug trial for PAH. Each patient had undergone cardiac catheterization and complete medical workup to exclude other significant disorders and to establish the diagnosis of PAH as the primary disorder. As a requirement of the study, which was approved by the Investigational Review Boards at each site, each patient had duplicate CPET performed, usually 1 day but as long as 3 days apart. Each paired set of CPET studies was repeated every 3 months in all patients for up to 15 months.

**Protocol**

At both sites, after familiarization with the CPET equipment, patients underwent an incremental CPET on a cycle ergometer under physician supervision, breathing through a mouthpiece with nose clipped. The protocols were identical at both sites: 3 min of sitting rest, 3 min of unloaded cycling; increasing work rate in a ramp fashion at a rate of 5 to 15 W/min (mostly 10 W/min) until symptom-limited, followed by 2 to 3 min of recovery.

**Equipment and Data Display**

At both sites, CPETs were performed using computer-controlled cycle ergometers with either a MedGraphics Cardiopulmonary Exercise System (Medical Graphics; St. Paul, MN) or a SensorMedics Vmax System (SensorMedics; Yorba Linda, CA). Calibration was done according to the instructions of the manufacturers. Gas exchange and cardiovascular data were retrieved electronically after exercise. Each exercise study was retrieved from its electronic file. The data were computer averaged for either 10-s or 15-s periods and formatted by one investigator into graphical and tabular displays at one of the sites so that the graphical displays were similar from the two sites. The data given to each grader consisted of nine-panel graphs on a single page, an enlarged $CO_2$ output ($V_{e\text{t}CO_2}$) vs $V_{O_2}$ graph with equal axes on a single page for optimal discrimination of the AT by the V-slope method, and tabular data on one to two pages.

**Data Evaluation**

Each study set of three to four pages was coded, randomized, given a new identity number, and copied so that evaluators receiving the printed data were blinded as to the patients’ identities, dates of the study, their cohort, or any investigator’s prior interpretations of their studies. For each test, six evaluators, four from one site and two from the other site, agreed to independently record their numerical interpretation of the data on preprinted blank forms, following specific instructions. Each evaluator received the guidelines shown in Table 1 to determine the seven exercise measurements listed. To obtain maximal stability, $Pe\text{t}CO_2$@AT and $Ve/\dot{V}CO_2$@AT values were averaged at and immediately after the AT was reached, while $Pe\text{t}O_2$@AT values were averaged immediately before and at the AT. Each evaluator in interpreting the AT ranged from 1 year to $>25$ years.

The written tabular data from each of the six evaluators were collected and keyed into a new database by a single investigator. On initial evaluation, when a single datum recorded by an evaluator was found to deviate strikingly from the mean of the data of other evaluators (frequency < 1%), the possibility of a transcription or calculation error was considered. In such a circumstance, that evaluator was notified that such an error may have occurred for that single datum; the evaluator was allowed to review and resubmit the datum for that single datum. There was no feedback given to evaluators regarding their performance during their evaluation of the study data.

**Statistical Analysis**

**Single Tests:** For the 228 exercise tests evaluated, the mean, SD, and coefficient of variation (CoV) for the group for each
variable were calculated. To compare the role of investigator/evaluator experience on the variability of AT measurements, paired t tests with Bonferroni corrections were used to compare the two most experienced, two intermediate experienced, and two least experienced evaluators.

**Paired Tests:** Using the above mean values and SDs for each variable for the 114 pairs of CPET studies, Pearson correlation coefficients (r) and CoV were calculated. Absolute and percentage differences between the first and second tests of each pair were summarized so that their mean and median values of the variable differences could be calculated. (For example, if the peak VO2 for tests 1 and 2 in one patient were 950 mL/min and 1,050 mL/min, respectively, the mean, SD, CoV, and absolute and percentage differences would be 1,000 mL/min, 70.7 mL/min, 7.1%, 100 mL/min, and 10%, respectively.) The group mean values for each variable for the first and second of each paired test were graphed using equal X-Y coordinates according to the Bland-Altman method.31 To assess the importance of the size of the measured value, we also calculated the r, slope, and intercept values as each variable increased. Analysis of variance was performed to ascertain whether or not differences in patient age or the number of repeated paired tests caused significant trends or differences in any measurement variability; p < 0.05 was considered significant.

To assess what magnitude of differences between serial tests might be clinically significant in a single patient with severe heart or lung disease, we considered that physiologic improvement might show consistent increases in peak VO2, peak O2P, AT, PETCO2@AT, and decreases in Ve/VCO2@AT and PETCO2@AT greater than those found in reproducibility of the paired tests reported here. Physiologic deterioration would show the opposite changes. Therefore, we averaged the consistent change of these six separate independent physiologic variables, all weighted equally, for each pair of tests (either improving or deteriorating), and sorted the 114 paired tests from minimal to maximal change.

### RESULTS

There was no morbidity associated with CPET; none of 228 exercise tests (114 pairs) were discarded. The patients ranged in age from 9 to 59 years and in peak VO2 from 24 to 69% of predicted.35 Thirty-five of the 42 patients in the study were female (83%). All but eight were Weber class C or D (Table 2).

For 228 single tests measured by the six blinded evaluators, the CoV of peak VO2, peak HR, peak O2P, and PETCO2@AT were <1%, for PETCO2@AT and Ve/VCO2@AT were approximately 2%, and for AT was 8.5% (Table 3). There were no significant differences or trends in variability with repeated testing.

In assessing the AT, experienced evaluators were...
in better agreement with each other than those with less experience. For the two least experienced evaluators, the median absolute difference in the AT was 60 mL/min; for the two evaluators with intermediate experience, it was 40 mL/min; and for the two most experienced evaluators, it was 20 mL/min (p < 0.0001 for these comparisons). These differences between evaluators were present whether or not the patients exhibited right-to-left shunting during exercise. Further, the presence of shunting did not significantly increase or decrease the differences between evaluators.

For 114 paired CPET studies, the absolute and percentage median differences between test 1 and test 2 measurements and CoV are shown in Table 4. The median differences and CoV both ranged from 1 to 7%.

Although the scales differ, the absolute differences between paired values were all relatively small when analyzed using the Bland-Altman plots (Fig 1, 2). There was no order effect. For most variables, the size of the absolute differences between paired values did not change appreciably as test values increased. The size of the absolute differences between paired tests increased significantly as their values increased only for peak $\dot{V}O_2$, peak $O_2$P, and $V\dot{E}/VCO_2@AT$ (Table 5). For peak $V\dot{O}_2$, the absolute difference between paired tests increased by 64 mL/min for each 1,000 mL/min increase in $V\dot{O}_2$, or 6.4%; for peak $O_2$P, the absolute difference increased by < 0.04 mL/L/beat per 1 mL/L/beat increase, or 4%. For $V\dot{E}/VCO_2@AT$, the significant r value (r = 0.446) became insignificant (r = 0.022) when its reciprocal times a constant, which equals the $PETCO_2@AT$, was calculated (Table 5, Fig 2). Other variable differences were insignificantly affected by a change in quantity.

In the assessment of clinically significant changes, 57 of the most highly reproducible pairs (50%) changed their average of the six physiologic measurements in a consistent direction by < 2.8%; 80 of the pairs (70%) changed by < 5.2%, and 103 of the pairs (90%) changed by < 10.0% (Fig 3). Neither patient nor the number of repetitions of pairs influenced variability between paired tests.

**DISCUSSION**

**Single-Test Interobserver Variability**

The evaluators in this study measured most CPET values very similarly in single tests (Table 3). We believe the strict rules given as a guideline (Table 1) were important in minimizing evaluator variability. The measurements of peak $V\dot{O}_2$, peak HR, peak $O_2$P, and $V\dot{E}/VCO_2@AT$ were especially consistent.

Measurements of the AT were more variable, even for more experienced evaluators, but the differences were relatively small even for the less experienced. In our prior experiences, AT differences between evaluators are reduced when a single evaluator has the opportunity to evaluate tests in the same patient serially or if a group of evaluators take the opportunity to discuss specific tests together. Neither of these tactics was used in this study. Subjectively, we previously had noted that AT measurement tended to be more difficult in patients who exhibit right-to-left shunting through an open foramen ovale during exercise, as the sudden increase in $VCO_2$, $PETO_2$, and ventilatory equivalents, accompanied by a decrease in $PETCO_2$, and usually decrease in oximetry saturation can be interpreted incorrectly as the onset of the AT, because the AT is also manifested by an increase in $VCO_2$, $R$, and minute ventilation/$V\dot{O}_2$. Thus, we were surprised to find that there was no significant increase in the variability of AT measurement in those patients that exhibited such shunting.

**Paired Tests**

The evaluation of paired tests in our study indicates that a single, well-performed test in a cooperative patient, even with severe or very severe exercise limitation, is valid. Despite slight differences in site equipment and technicians, time of day, patient

---

**Table 3—Mean and CoV of Seven Exercise Variables for 228 Single Tests as “Blindly” Determined by Six Evaluators**

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Mean</th>
<th>CoV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Peak V\dot{O}_2$, mL/min</td>
<td>746</td>
<td>0.6</td>
</tr>
<tr>
<td>$Peak HR$, beats/min</td>
<td>143</td>
<td>0.5</td>
</tr>
<tr>
<td>$Peak O_2$P, mL/L/beat</td>
<td>5.3</td>
<td>1.0</td>
</tr>
<tr>
<td>AT, mL/min</td>
<td>549</td>
<td>8.5</td>
</tr>
<tr>
<td>$PETO_2@AT$, mm Hg</td>
<td>121</td>
<td>1.0</td>
</tr>
<tr>
<td>$PETCO_2@AT$, mm Hg</td>
<td>27</td>
<td>2.2</td>
</tr>
<tr>
<td>$V\dot{E}/VCO_2@AT$</td>
<td>53</td>
<td>2.1</td>
</tr>
</tbody>
</table>

**Table 4—Median Absolute Differences and CoV in 114 Paired Tests**

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Absolute</th>
<th>%</th>
<th>CoV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Peak V\dot{O}_2$, mL/min</td>
<td>46</td>
<td>6.9</td>
<td>5.8</td>
</tr>
<tr>
<td>$Peak HR$, beats/min</td>
<td>5.7</td>
<td>4.0</td>
<td>3.3</td>
</tr>
<tr>
<td>$Peak O_2$P, mL/L/beat</td>
<td>0.30</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>AT, mL/min</td>
<td>36</td>
<td>6.4</td>
<td>6.5</td>
</tr>
<tr>
<td>$PETO_2@AT$, mm Hg</td>
<td>1.3</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>$PETCO_2@AT$, mm Hg</td>
<td>0.8</td>
<td>3.3</td>
<td>2.8</td>
</tr>
<tr>
<td>$V\dot{E}/VCO_2@AT$</td>
<td>1.6</td>
<td>3.8</td>
<td>3.3</td>
</tr>
</tbody>
</table>
Figure 1. Measurements of four key exercise variables in 114 paired tests in 42 patients with PAH. Upper left: peak VO\(_2\); upper right: peak HR; lower left: peak O\(_2\)P; and lower right: AT. Each quadrant depicts test 1 vs test 2 values with the line of identity in the upper panel plus a Bland-Altman plot immediately below. bpm = beats per minute.
Figure 2. Measurements near the AT in 114 paired tests in 42 patients with PAH. Upper left: $P_{ETO_2}@AT$; upper right: $P_{ETCO_2}@AT$; lower left: $VE/VCO_2@AT$; and lower right: $PECO_2@AT$. Each quadrant depicts test 1 vs test 2 values with the line of identity in the upper panel plus a Bland-Altman plot immediately below.
mood, intercurrent patient illnesses, or fluctuations in disease intensity, the measured values of a second test performed 1 to 3 days later usually differed only minimally from the first test. In fact, the median differences for paired tests for all variables ranged from only 1 to 7% for peak $\dot{V}O_2$, peak HR, peak $O_2P$, $AT$, and $VE/VECO_2@AT$ (Table 4). The $AT$ reproducibility may be especially important since some clinicians and investigators are reticent to have their patients exercise to their maximal tolerance for measurement of peak $\dot{V}O_2$, peak HR, and peak $O_2P$, and thus may choose to use the $AT$ as a surrogate for maximal aerobic capacity.

Measurement variability was not increased at either age extreme. Also, the variability between paired tests did not significantly change from the

### Table 5—Correlation Coefficients Between Paired Exercise Tests Plus Changes in Absolute Differences as Variable Values Increased

<table>
<thead>
<tr>
<th>Variables</th>
<th>Paired Test $r$ Value</th>
<th>Change in Absolute Differences as Variable Values Increase $r$ Value</th>
<th>Slope</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\dot{V}O_2$, mL/min</td>
<td>0.964*</td>
<td>0.374*</td>
<td>0.064</td>
<td>13.4</td>
</tr>
<tr>
<td>Peak HR, beats/min</td>
<td>0.923*</td>
<td>-0.059</td>
<td>-0.016</td>
<td>9.0</td>
</tr>
<tr>
<td>Peak $O_2P$, mL/beat</td>
<td>0.966*</td>
<td>0.200</td>
<td>0.036</td>
<td>0.29</td>
</tr>
<tr>
<td>$AT$, mL/min</td>
<td>0.902*</td>
<td>0.084</td>
<td>0.026</td>
<td>36.3</td>
</tr>
<tr>
<td>$PETO_2@AT$, mm Hg</td>
<td>0.949*</td>
<td>-0.076</td>
<td>-0.015</td>
<td>3.5</td>
</tr>
<tr>
<td>$PETCO_2@AT$, mm Hg</td>
<td>0.976*</td>
<td>-0.025</td>
<td>-0.003</td>
<td>1.2</td>
</tr>
<tr>
<td>$VE/VECO_2@AT$</td>
<td>0.970*</td>
<td>0.446*</td>
<td>0.081</td>
<td>-1.8</td>
</tr>
<tr>
<td>$PECO_2@AT$, mm Hg</td>
<td>0.977*</td>
<td>0.022</td>
<td>0.031</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

*p < 0.001 for Pearson correlation coefficient and for difference in slope from zero.

$p < 0.05.$

![Figure 3](http://journal.publications.chestnet.org/pdaccess.ashx?url=/data/journals/chest/22015/ on 06/25/2017)

**Figure 3.** Magnitude and frequency of change between 114 paired tests. The x-axis is the mean percentage change in a consistent direction for peak $\dot{V}O_2$, peak $O_2P$, $AT$, $VE/VECO_2@AT$, $PETCO_2@AT$, and $PETO_2@AT$. The y-axis indicates the individual and cumulative incidence of these changes. The dashed lines indicate the 2.5%, 5%, and 10% average change for approximately 50%, 70%, and 90% of the paired tests, respectively.
first pair to the second, third, fourth, fifth, or sixth pair of tests. Quantitatively, the modest increase in absolute differences between paired tests for peak \( \dot{V}O_2 \) and peak \( O_2 \) pulse at higher values (Fig 2, Table 5) is not surprising. In fact, the lower absolute differences in patients with more severe disorders are salutary.

The variability of \( V_{E}/V_{CO_2}@AT \)-paired differences significantly increased as their values increased from 30 to 90 U. In contrast, the variability of \( P_{ACO_2}@AT \)-paired differences remained stable from 30 to 10 mm Hg (Fig 2, Table 5). The \( P_{ACO_2} \) and \( V_{E}/V_{CO_2} \) values are reciprocal. Thus, as the \( P_{ACO_2}@AT \) values declined and ventilatory efficiency decreased, the variability between paired tests for \( V_{E}/V_{CO_2}@AT \) significantly increased.

Factors Not Assessed

Our study did not assess the reproducibility of peak \( V_{CO_2} \), peak respiratory exchange ratio, or peak ventilation because these values are somewhat dependent on recent diet and substrate availability. We did not assess peak work rate, as it is more variable and also of lesser clinical utility than peak \( \dot{V}O_2 \). Nevertheless, other investigators examining highly trained cyclists and a large number of normal adults found test-retest correlation coefficients > 0.86 for peak \( \dot{V}O_2 \), peak \( V_{CO_2} \), peak ventilation, peak HR, and peak work rate. We did not assess \( V_{E}/V_{O_2} \) or \( V_{E}/V_{CO_2} \) at the peak work rate, because these values are dependent on the severity of acidemia during exercise and thus do not reflect maximal gas exchange efficiency. It would have been useful for us to have also assessed the nadir \( V_{E}/V_{CO_2} \) after the AT or slope of minute ventilation vs \( V_{CO_2} \), as these have variability similar to that of \( V_{E}/V_{CO_2}@AT \) in a normal population.

Clinical Relevance

The high reproducibility of the key physiologic measurements in our study compares favorably with other studies involving patients with heart, lung, musculoskeletal, or renal disease with less severe exercise impairment. For example, in 11 patients with end-stage renal disease and moderate exercise impairment (mean peak \( \dot{V}O_2 \) of 1,500 mL/min), Koufaki et al reported CoVs for peak \( \dot{V}O_2 \), peak HR, and peak \( O_2 \)P of 4.7%, 5.9%, and 6.0%, respectively. In an exacting study of 17 patients with mild heart disease, who were hospitalized, fasted, and tested at the same time each day, Lehmann and Kolling reported correlation coefficients of 0.996 for peak \( \dot{V}O_2 \), 0.928 for peak HR, 0.991 for peak \( O_2 \)P, and 0.995 for \( V_{E}/V_{CO_2}@AT \). Their very high correlation coefficients can be partially attributed to the broad range of peak \( \dot{V}O_2 \) in their patients (803 to 3,927 mL/min). In 16 cardiac patients with moderate impairment (Weber class A for 6 patients, Weber class B for 5 patients, and Weber class C for 5 patients, mean peak \( \dot{V}O_2 \) of 26.5 mL/min/kg, 17.4 mL/min/kg, and 13.8 mL/min/kg, respectively) tested several times over a period of 3 to 22 months, Janicki et al found CoVs of 5.7%, 4.4%, and 9.2% for peak \( \dot{V}O_2 \), peak HR, and AT, respectively. In duplicate tests in 11 cardiac patients (4 patients were New York Heart Association class II, and 7 patients were class III), Meyer et al in patients with mean peak \( \dot{V}O_2 \) of 13.9 mL/min/kg, found CoVs for peak \( \dot{V}O_2 \), peak HR, and peak \( O_2 \)P of 4.1%, 1.4%, and 4.4%, respectively.

Several studies of CPET reproducibility in patients with obstructive or restrictive lung disease have also been reported. In triplicate tests in nine adult patients with cystic fibrosis (mean predicted peak \( \dot{V}O_2 \) of 62%), McKone et al found a CoVs for peak \( \dot{V}O_2 \) and peak HR were 6.9% and 3.0%, respectively. In triplicate tests in six patients with restrictive lung disease and predicted peak \( \dot{V}O_2 \) of 58 ± 10%, Marcinuk et al found a CoVs of 5.3% for peak \( \dot{V}O_2 \) and 4.0% for peak HR. In 11 patients with obstrusive lung disease, familiar with CPET and mean peak \( \dot{V}O_2 \) of 1.3 L/min, Brown et al found a CoVs of 7%. In 11 patients with mild obstructive lung disease (peak \( \dot{V}O_2 \) of 1.95 L/min), Cox et al reported CoVs for peak \( \dot{V}O_2 \) and peak HR of 3% and 4%, respectively. In duplicate tests in 56 patients with COPD with a mean predicted peak \( \dot{V}O_2 \) of 65%, Covey et al found correlation coefficients and CoVs of 0.97 and 4.5%, and 0.92 and 3.0%, respectively, for peak \( \dot{V}O_2 \) and peak HR.

The patients with PAH in the current study appear to be more impaired than those in any previously reported CPET reproducibility study of patients with either cardiovascular or pulmonary diseases, as their mean ± SD percentage of predicted peak \( \dot{V}O_2 \) was only 43 ± 10% and their Weber classes were usually severe or very severe (81% C or D). Yet, the reproducibility of their key CPET measurements remained high. Most of the studies cited above, as did ours, used cycle ergometry, which has the advantage of being under patient control and also of allowing the calculation of the \( \dot{V}O_2/ \)work rate relationship. However, for the measurements that were addressed in this study, there is no reason to expect that they would not be similarly reproducible using treadmill ergometry.

Thus we conclude that single, well-conducted gas exchange exercise tests can safely and reliably measure key and discriminating exercise variables, including peak \( \dot{V}O_2 \), in moderately impaired to very severely impaired children and adults. The median
difference in peak VO2 was 7%, with lesser differences in the other measurements, all relatively independent of the absolute value. Thus, taken alone, a change in peak VO2 of 8 to 10% might be used as a cut off in deciding that there is a significant change in exercise tolerance. However, using the combination of peak VO2, peak O2P, AT, PETCO2@AT, Ve/VO2@AT, and PETO2@AT from one test to the next the average consistent change would be < 3% in half of all patients, < 5% in 7 in 10 of all patients, and < 10% in 9 in 10 of patients. To increase uniformity of reading and interpreting aerobic function and ventilatory efficiency using CPET, we recommend that other practitioners consider using the guidelines given in Table 1.

ACKNOWLEDGMENT. The authors thank Dr. Ron Oudiz for management and referral of many of the patients, and for his assistance in study design and execution and manuscript revisions.

REFERENCES