New graphical method for evaluating gas exchange in congestive heart failure

ROBERT L. BARD and JOHN M. NICKLAS

University of Michigan Medical Center, Congestive Heart Failure and Transplant Program, Ann Arbor, MI

ABSTRACT

BARD, R. L. and J. M. NICKLAS. New graphical method for evaluating gas exchange in congestive heart failure. Med. Sci. Sports Exerc., Vol. 32, No. 4, pp. 870–876, 2000. Purpose: The purpose of this study was to determine the usefulness of a new graphical method for evaluating gas exchange (GE), as applied to three common averaging techniques used during rest, exercise, and recovery in patients with congestive heart failure (CHF). Methods: Fifty patients’ resting, exercise, and recovery GE graphs were evaluated using a new graphical method to determine oxygen consumption (V\(\text{O}_2\)), ventilation (VE), and respiratory exchange ratio (RER). Results: There were no significant differences of repeated measures between trained and untrained investigators. The graphical method was evaluated by repeated measures using the same investigator, an additional investigator, and an untrained student. A 1-min rolling time average (RTA) updated every 20 s was used as the reference to the graphical method and the metabolic cart’s computer analysis. Breath-by-breath data were evaluated using three averaging techniques, average 5 of 7 breaths (5/7), average of 8 breaths (AVG 8), and 30-s average (30-s). Differences between the computer and the graphical method were assessed by two-sided t-tests; ANOVA was used between the three investigators and also between the three averaging techniques. Results: There were no significant differences of repeated measures between trained and untrained investigators. The graphical method results were significantly different than the computer results. The only difference with RTA involved computer resting values, which were also different than the graphical method. Conclusion: In conclusion, the graphical method may be used to standardize GE evaluation because it displays excellent intrainvestigator precision and good interinvestigator precision between experienced and inexperienced investigators. Averaging techniques less than 30 s have greater bias when using computer-derived values, although when incorporating the graphical method the averaging technique chosen has little influence on all measures. Key Words: CARDIOPULMONARY EXERCISE TESTING, CARDIOMYOPATHY, OXYGEN CONSUMPTION, METABOLISM
Periodic breathing is known to be associated with increased mortality risk (1,6,9,10,19) and elevated pulmonary capillary wedge pressures (1,7,19). One author has expressed the need for standardization of sampling intervals for exercise testing (23) and has since recommended using 30-s averages printed every 10 s (21,22). This “rolling time average” (RTA) is unique and is not commonly available on automated systems. In the review by Howley et al. (11), a 30-s time interval is suggested for exercise testing.

The purpose of this study was to determine the usefulness of a new graphical method for evaluating GE as applied to three common averaging techniques during rest, exercise, and recovery in patients with CHF.

**METHODS**

This protocol was designed to mirror GE interpretation in clinical practice. It is the University of Michigan Heart Failure Program’s practice to determine numerical values of each test from the printed graphs of data. This procedure is referred to as the “graphical method”. This method allows clinicians to determine values directly and to account for outliers themselves, without the bias of the computer’s analysis. This was a retrospective study where the GE of 50 consecutive patients was evaluated during three different stages of the test: rest, exercise, and recovery. An exercise physiologist who has used this technique over a 2-yr period...
in transplant evaluations and research protocols tested all patients and evaluated each graph. Before each test, comprehensive written informed consent was obtained from each patient, and the metabolic cart was calibrated using the manufacturer’s standard procedures (17).

Three different averaging techniques were incorporated: rolling middle five of seven breaths (5/7), rolling average of eight breaths (AVG 8), and time average of 30 s (30-s). The 5/7 technique uses the current breath and the six preceding it, omits the minimum and maximum values of the product of VO₂ and VE, and averages the remaining five. The AVG 8 technique uses the current breath and averages it with the seven preceding breaths, whereas the 30-s technique averages all of the data collected over a 30-s period. All patients were tested using a Medical Graphics CPX/D metabolic cart with a disposable mouthpiece and pneumotach (Medical Graphics Corporation, St. Paul, MN) during all stages of the test.

V̇O₂ and V̇E were analyzed for all stages, with ventilatory threshold (VT) and respiratory exchange ratio (RER) included in the exercise stage. As in clinical practice, the VT was evaluated from the display of data on the computer screen using the V-Slope method (3), which is built into the Medical Graphics software. This method has been reported as being superior in evaluating patients with CHF (24), and the computer software allows the user to toggle between each data point.

Graphical method. The GE variables of each stage of the test were graphed against time, and were printed to the limits of the operating system (approximately one page each), for each of the averaging techniques. Each graph was evaluated using a fine mechanical pencil and a transparent graphing triangle. Figure 1 displays one subject’s test using the graphical method.

Rest. After approximately 10 min in Fowler’s position (supine with torso elevated 30 degrees on an exam table) during preparation for exercise, GE was collected for three consecutive minutes. The graphical method resting values were defined as the average of the last two of the 3 min of data collected during rest. The graphing triangle was placed at the estimated mean value of each variable over the last 2 min of rest, and a line was drawn to its corresponding axis to determine its value (Fig. 1).

Exercise. The graphical method peak values were defined as the highest linear value achieved during incrementally ramped treadmill exercise. The graphing triangle was placed along the estimated mean of the oscillating exercise data during the terminal portion of the exercise test, and a line of best fit was drawn. A line parallel to the horizontal axis was drawn from the vertical axis to the point where the data line (connection of data points) and the line of best fit intersected to determine the value of each variable (Fig. 1).

Recovery. The graphical method recovery values were defined as the lowest value achieved after 5 min of recovery: 2 min of active recovery at 1 mph and 3 min of inactive recovery in Fowler’s position. A line parallel to the horizontal axis was drawn from this lowest point to the vertical axis to the value of each variable (Fig. 1).

Precision. Precision of the graphical method was evaluated using VO₂ from each of the three stages from 10 tests.
All investigators were blinded to the subject and averaging technique and were provided with their own copy of the graphs. Intraintigator precision involved repeated measures of the same tests, with a 12-wk period between repeated measures. Interinvestigator precision involved three investigators: the initial investigator (investigator No. 1), a cardiologist in the heart failure program who uses this technique in practice (investigator No. 2), and a student who was inexperienced in the technique (investigator No. 3). Investigator No. 3 was provided with only the written instructions from this manuscript to evaluate each graph.

**Computer analysis.** The Medical Graphics Breeze 3.02a software (Medical Graphics Corporation, St. Paul, MN) was used on the metabolic cart and provided the resting and exercise computer-derived values. All of the patients’ computer-derived exercise and rest variables were recorded by 5/7, AVG 8, and 30-s averaging techniques for comparison to the graphical method. The computer software package does not evaluate recovery data.

**Reference.** The graphical method and computer analysis were compared with an RTA as an objective value for resting and exercise data. The RTA was evaluated over the last minute of data collection, which was updated every 20 s. GE variables were printed from the metabolic cart breath by breath, and the RTA was computed by hand.

**Statistical analysis.** One-way ANOVA was performed between different averaging techniques of each variable and between investigators with a level of significance of 0.05. Two-sided t-tests were performed between the computer and graphical method of each averaging technique for each variable. Data are presented in figures using Bland and Altman procedures (5). All statistical procedures were performed using Microsoft Excel 7.0 software (Microsoft Corporation, Seattle, WA).

**RESULTS**

Fifty patients’ metabolic data were evaluated. The patients’ mean characteristics were as follows: 49.1 ± 9.54 yr old, 11 women and 39 men, 77.3 ± 1.8 kg, 173.9 ± 9.8 cm tall, most recent left ventricular ejection fraction of 24.6 ± 11.6% retrieved from patient records, and a peak VO₂ of 17.9 ± 4.5 mL·kg⁻¹·min⁻¹.

Figure 2 displays the comparisons between the computer analysis, graphical method, and RTA for the resting data. There were no statistically significant differences between RTA and the graphical method for VO₂, whereas resting VO₂ was statistically different between the computer and both the graphical method and RTA. Resting VE was statistically different for both the graphical method and the computer in comparison to the RTA, whereas the computer and graphical method VE were not significantly different.

Figure 3 displays the comparisons between the computer analysis, graphical method, and RTA for the exercise data. There were no statistically significant differences between RTA and either the graphical method or the computer results, whereas there were differences between the computer and the graphical method for exercise VO₂, VE, and RER.

The graphical method revealed no statistically significant differences between trained and untrained investigators.
(Figs. 4, 5, and 6), or between different averaging techniques (Table 1) for rest, exercise, or recovery. Three of the 50 patients did not achieve VT by any of the techniques, whereas one patient clearly attained VT using both 5/7 and AVG 8, but VT was indeterminate by 30-s. Table 2 shows the \( P \) values from the two-sided \( t \)-tests between the computer and graphical methods for the resting and exercise values.

**DISCUSSION**

**Graphical method.** The graphical method proved to be helpful in the evaluation of GE in patients with CHF. The computer-derived results were statistically different and proved unreliable compared with the graphical method, particularly at rest. Although the “true” value is unknown, Figure 1 displays one example when the computer reported unrealistic values that were unsubstantiated by the data. The computer was also in poor agreement with the graphical method and the RTA (Fig. 2). The reasons for this discrepancy are unknown, but are likely technical error from the computer software program. There were three cases where the computer resting \( \text{VO}_2 \) values exceeded the reported VT values, and eight cases where the \( \text{VO}_2 \) or VE exceeded twice the graphical method. The precision of the graphical method was limited by the variability within patients. Many patients exhibited artifact and periodic breathing where the graphs were quite variable, making it difficult to consistently apply exact numeric data to the data.

The exercise stage of the evaluation can be particularly difficult for the computer to evaluate because of abnormal ventilation and nonphysiologic artifact (14). Periodic breathing is a common ventilatory abnormality in exercising patients with CHF (4), which tends to be more prevalent in the types of patients referred for transplant evaluation who have more severe heart failure (1,6,9,10,19). Additionally, GE artifact can be exaggerated when the workloads are altered by factors such as use of the treadmill handrails.

These factors demonstrate the importance of smoothing data in a manner where the sampling interval is large enough to be indicative of the patient’s physiology. The use of small averaging techniques allows a short time interval to provide a minute value, which may be multiplying measurement errors. This tendency is visible in Figure 3, where the bias for all measures is greatest for the sampling intervals that are less than 30 s. The computer uses a systematic search for the highest \( \text{VO}_2 \) attained, which may be an outlying spike from the graph, and reports it as the patient’s peak minute value. When using larger time intervals, this error is reduced but not removed.

The VE and RER comparisons in Figure 3 are not reliable because of the manner in which the computer selects peak values. In the computer’s analysis the peak \( \text{VO}_2 \) is noted,

![Graphs showing VO2 values](image)

**Table 1.** Mean resting, exercise, and recovery values from the graphical method (\( N = 50 \)).

<table>
<thead>
<tr>
<th></th>
<th>( \text{VO}_2 ) (ml·kg(^{-1})·min(^{-1}))</th>
<th>VE (L·min(^{-1}))</th>
<th>VT (ml·kg(^{-1})·min(^{-1}))</th>
<th>RER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph 5/7</td>
<td>2.77 ± 0.787</td>
<td>9.36 ± 2.120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph AVG 8</td>
<td>2.73 ± 0.797</td>
<td>9.16 ± 2.355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph 30 sec</td>
<td>2.90 ± 0.744</td>
<td>9.00 ± 2.333</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exercise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph 5/7</td>
<td>17.85 ± 4.52</td>
<td>62.37 ± 19.49</td>
<td>11.27 ± 2.58 (( N = 47 ))</td>
<td>1.122 ± 0.077</td>
</tr>
<tr>
<td>Graph AVG 8</td>
<td>17.86 ± 4.61</td>
<td>61.15 ± 19.27</td>
<td>11.30 ± 2.59 (( N = 47 ))</td>
<td>1.131 ± 0.083</td>
</tr>
<tr>
<td>Graph 30 sec</td>
<td>17.50 ± 4.47</td>
<td>59.46 ± 18.70</td>
<td>11.54 ± 2.94 (( N = 46 ))</td>
<td>1.114 ± 0.081</td>
</tr>
<tr>
<td><strong>Recovery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph 5/7</td>
<td>4.05 ± 1.33</td>
<td>17.38 ± 5.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph AVG 8</td>
<td>4.21 ± 1.34</td>
<td>17.94 ± 5.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph 30 sec</td>
<td>4.16 ± 1.29</td>
<td>17.55 ± 5.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values are mean ± SD.
TABLE 2. P values from two-sided t-tests between computer and graphical methods for resting and exercise values (N = 50).

<table>
<thead>
<tr>
<th></th>
<th>Computer</th>
<th>Graphical</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VO₂</strong></td>
<td>0.032*</td>
<td>0.225</td>
</tr>
<tr>
<td><strong>VE</strong></td>
<td>0.015*</td>
<td>0.095</td>
</tr>
<tr>
<td><strong>RER</strong></td>
<td>0.008*</td>
<td>0.000*</td>
</tr>
<tr>
<td>30 sec</td>
<td>0.040*</td>
<td>0.026*</td>
</tr>
</tbody>
</table>

* Statistically significant.

and the RER and VE associated with this VO₂ are reported as the “peak” values. These values are not peaks, but rather are those associated with what the computer believes is the peak VO₂. In Figure 3, these values are being compared with the true peak values as assessed using the graphical method. The computer analysis has the potential to lead to error and inconsistencies when values are reported that may not be representative of the patient’s true capacity. A patient may truly achieve an RER ≥ 1.10, but a lower RER may be reported if it was associated with an outlying VO₂. The graphical method allows the clinician to select all GE variables independently, accounting for artifact to determine a more representative value of the patient’s true capacity.

The graphical method was superior in the exercise evaluations, as in the resting data. Figure 3 displays less bias using the graphical method than the computer analysis compared with the RTA for peak VO₂. The bias was always greatest with the computer results, exceeding 4 mL·kg⁻¹·min⁻¹ for both of the shorter intervals. The range of the computer 30-s bias was more acceptable (−1.4 to 2.0 mL·kg⁻¹·min⁻¹) but was still greater than the range of the bias of the graphical method (−1.4 to 0.4 mL·kg⁻¹·min⁻¹).

Clinicians base important clinical judgments on a test that is lacking standardization throughout the medical community (11). Many medical centers use peak VO₂ as a strict transplant criteria; if peak VO₂ is less than 14.0 mL·kg⁻¹·min⁻¹ with a peak RER ≥ 1.10, patients can be listed for transplantation (8), whereas cardiac transplantation may be deferred with a peak VO₂ > 14.0 mL·kg⁻¹·min⁻¹ (20). In this scenario, standard procedures are of great importance in evaluating peak VO₂ and for evaluating RER independent of the peak VO₂ to insure the proper allocation of donor hearts. The patient whose test is displayed in Figure 1 would be placed on the transplant list using the graphical technique, yet would not be listed if the computer-derived value were used. The intrainvestigator precision of peak VO₂ measurement used the graphical method was excellent, which can help to standardize transplant evaluations at individual medical centers. The precision of the measurement decreased as other investigators were added, yet the bias was typically within 1.0 mL·kg⁻¹·min⁻¹ (Fig. 5).

GE measurements can have profound economic impacts on health care providers, insurance companies, and the social security administration because of the payment of disability benefits. Patients who have peak VO₂ values under 15 mL·kg⁻¹·min⁻¹ can be declared unfit or incapacitated for employment, and patients may be employed if their peak VO₂ is in the range of 15–24 mL·kg⁻¹·min⁻¹ (26). At times the bias between both the computer and the RTA and the computer and the graphical method exceeded 4 mL·kg⁻¹·min⁻¹, which would easily misclassify many patients. The lack of standardization of oxygen consumption testing is potentially misclassifying many patients and not allowing for consistent evaluation.

The graphical method may be used effectively in clinical practice, in research studies, and in large clinical trials to standardize reporting of results because it was shown to be precise regardless of the experience of the investigator (Figs. 4, 5, and 6).

**Averaging techniques.** The agreement between the three techniques was consistent with previous studies that found shorter averaging intervals to elicit higher values (12,16). This trend was displayed in both the graphical method and the computer selections. There were no significant differences for the exercise data, but VT was difficult to pick using the 30-s technique because fewer data points limit the investigator’s ability to demarcate a threshold. This limitation is particularly evident during short exercise tests, which are common in the evaluation of end-stage heart failure patients.

**CONCLUSION**

In conclusion, the graphical method may be used to standardize GE evaluation because it displays excellent intrainvestigator precision and good interinvestigator precision between experienced and inexperienced investigators. Averaging techniques less than 30 s have greater bias when using computer-derived values, although when incorporating the graphical method, the averaging technique chosen has little influence on all measures.

Address for correspondence: Robert L. Bard, University of Michigan Medical Center, 1500 E. Medical Center Drive, L3119 Womens, Box 0273, Ann Arbor, MI 48109-0273. E-mail: bbard@umich.edu.

**REFERENCES**


