Determination of Gas Exchange Threshold by Nonparametric Regression

ULYSSES J. MAGALANG and BRYDON J. B. GRANT

Department of Medicine, State University of New York at Buffalo, Buffalo, New York

The gas exchange threshold (GET) has been used an an index of anaerobic threshold because it can be measured noninvasively. GET is estimated from a breakpoint in breath by breath values of carbon dioxide uptake ($\dot{V}CO_2$) and oxygen uptake ($\dot{V}O_2$) obtained during a progressive exercise test. Three methods of estimating GET were evaluated: (*1*) the original V slope method (OVS) using two adjoining standard linear regressions, (*2*) the modified V slope method (MVS) where the breakpoint is detected by visual inspection, and (*3*) a new method that we developed with nonparametric regression (NPM) using cubic splines. Simulated data were used because the existence of a breakpoint is known with certainty. Detection accuracy for OVS and MVS never exceeded 63% because of a low specificity. The detection accuracy of NPM ranged between 50 and 89% depending on the amount of noise and abruptness of the threshold, and exceeded that of OVS and MVS at low levels of noise. NPM was significantly more accurate ($p < 0.05$) than OVS and MVS for detecting GET except with high levels of noise. Both NPM and OVS have similar degrees of numerical accuracy and are superior to the currently used MVS method in this respect. All three methods gave similar results on 20 exercise tests. We conclude from the simulated data that NPM is more accurate than OVS and MVS at detecting GET. NPM can be applied to human data and it provides results that are consistent with OVS and MVS. **Magalang UJ, Grant BJB. Determination of gas exchange threshold by nonparametric regression. Am J Respir Crit Care Med 1995;151:98-106.**

Wasserman and McIlroy (1) introduced the concept of anaerobic threshold (AT). The AT has been defined as the work rate at which anaerobic metabolism supplements aerobic mechanisms during progressive exercise (1). Invasive methods of AT detection involve measurements of the rise in blood lactate concentration (lactate threshold) or the fall in standard bicarbonate concentration (lactic acidosis threshold) (2).

Gas exchange measurements during exercise have been widely utilized as an index of AT because they can be determined noninvasively. The onset of metabolic acidosis is most commonly detected by a breakpoint in the breath by breath values of carbon dioxide output ($\dot{V}CO_2$) plotted against oxygen uptake ($\dot{V}O_2$) caused by an increase in $\dot{V}CO_2$ from bicarbonate buffering of lactic acid. The breakpoint has been termed the gas exchange threshold (GET) (2). Beaver and coworkers (3) described the original V slope method (OVS) of detecting GET using a computer program that generates two adjoining linear regressions. This method was believed to be superior to earlier approaches which were dependent on the regularity of breathing pattern and respiratory chemosensitivity.

Subsequently, Sue and colleagues (4) described a modified V slope method (MVS). The breakpoint was detected by visual inspection of the plots of $\dot{V}CO_2$ as a function of $\dot{V}O_2$. Although it appears that MVS has become more popular than OVS, there is no current agreement on which method to utilize in the detection of GET (2, 5–7) partly because of lack of studies directly comparing these two methods. The problem is complicated by the fact that no true reference standard exists to determine the test characteristics of these two methods. Use of lactate and standard bicarbonate measurements as reference tests can be misleading because they are themselves subject to error and are limited by the frequency by which arterial blood samples are obtained.

GET/AT has been used clinically in the differential diagnosis of exercise limitation (8), in the evaluation of patients with cardiovascular disease (9), in the pulmonary rehabilitation of patients with chronic obstructive lung disease (10), and most recently in the preoperative evaluation of elderly patients (7). While these data do provide support for an empiric relation between GET/AT and clinical status, the mechanism remains controversial (11–13). There is no uniform agreement that a threshold has been demonstrated convincingly statistically. Some have argued that the increase of arterial blood lactate with exercise can be fitted as well with exponential functions as with two linear regressions (11). Furthermore, the sensitivity, specificity, and interobserver agreement of MVS have been reported to be poor by some investigators (14). OVS circumvents the problem of interobserver variation because it is a computer-based method and therefore requires predetermined criteria, but its sensitivity and specificity have not been determined. On the other hand, OVS may be vulnerable to inaccuracies because the two linear regressions are imposed on the data regardless to the actual form.

The standard linear regression analysis assumes that the variance is uniformly scattered around each linear regression over the range of measurement. As a result, an outlying data point can

(*Received in original form January 3, 1994 and in revised form June 13, 1994*)

Supported in part by NIH HL-41011.

Correspondence and requests for reprints should be addressed to B. J. B. Grant, M.D., ECMC (Pulmonary Lab), 462 Grider Street, Buffalo, NY 14215.

Am J Respir Crit Care Med Vol 151. pp 98–106, 1995

have undue influence on the analysis. To circumvent this problem, extensive data preparation is required. Some of the procedures used in this process, such as moving averages, may blur the location of the breakpoint. An alternative approach is to use nonparametric regression which fits the data with cubic splines and does not assume uniformity of the variance of the data around the fitted regression line (15).

This study compares the test characteristics of the original and the modified V slope methods. In addition, a new method for calculating GET is described by the same criteria as OVS but using nonparametric regression (NPM). NPM not only provides a means of testing the existence of a threshold, but also provides an objective method for its detection. Most of these comparisons were made on simulated data sets. Simulated data are the only circumstance when the existence and value of a threshold can be known with absolute certainty.

METHODS

Definition of Terms

In this report, the "gas exchange threshold" is defined as an abrupt change in the first derivative of the relation between $\dot{V}CO_2$ and $\dot{V}O_2$ ($d\dot{V}CO_2/d\dot{V}O_2$). The breakpoint is detected from a local maximum in the plot of the second derivative ($d^2\dot{V}CO_2/d\dot{V}O_2^2$). If the data that describe this relation were continuous, a threshold would be defined as a discontinuity in the first derivative and the plot of the second derivative against $\dot{V}O_2$ would show a sharp spike (Figure 1).

"Detection accuracy" is the correctness of the response as to whether or not a threshold exists. A high detection accuracy requires both high sensitivity and high specificity. "Sensitivity" is defined as the proportion of subjects with a threshold that have a positive result (true positive rate). "Specificity" is defined as the proportion of subjects without a threshold that have a negative result (true negative rate). "Numerical accuracy" is the closeness of an estimate to the true value. An estimate is accepted as being accurate if the true value is within the 95% confidence limits of the estimate. Numerical accuracy is of little value unless it is accompanied by reasonable precision. "Precision" is defined as the degree to which multiple estimates of the threshold agree with each other. It is expressed quantitatively as the normalized standard error which is the standard error of the mean divided by the true value of the mean.

Simulated Data

Each data set was generated with and without a breakpoint and was comprised of 200 pairs of values for $\dot{V}CO_2$ and $\dot{V}O_2$. Data with a GET were generated from two contiguous linear relations. The lower segment started at $\dot{V}CO_2$ of 240 ml/min and $\dot{V}O_2$ of 300 ml/min. Each breath increased by 10 ml/min for $\dot{V}O_2$ with a slope of unity to $\dot{V}O_2$ of 1,300 ml/min. $\dot{V}O_2$ then increased by the same increments with a slope of either 1.1, 1.25, 1.5, or 1.75 up to $\dot{V}O_2$ maximum of 2,300 ml/min. Noise was introduced into the data by adding to each $\dot{V}CO_2$ and $\dot{V}O_2$ value from a set of 400 random numbers. The random numbers had a mean of zero and standard deviation of unity. The random component of each number was corrected to produce noise levels of 5, 25, 75, and 125 ml/min by multiplication. Thus, 16 conditions were produced: four levels of noise at four different ratios of the slopes of the two linear relations. For each condition, 100 data sets were generated.

Similar data sets were generated for each condition without a GET by using a Lagrangian quadratic expression (16) that passes through the initial point ($\dot{V}CO_2$ of 240 ml/min and $\dot{V}O_2$ of 300 ml/min), through GET ($\dot{V}CO_2$ of 1,240 ml/min and $\dot{V}O_2$ of 1,300 ml/min) and ends at $\dot{V}O_2$ maximum of 2,300 ml/min. Examples of four data sets are shown in Figure 2.

Human Data

We studied five normal subjects and 10 consecutive patients referred to the Pulmonary Function Laboratory of the Erie County Medical Center for exercise testing (Sensormedics 2900; Yorba Linda, CA). Clinical details are provided in Table 1. The Institutional Review Committee approved the use of these data for this study. Each normal subject exercised on a cycle ergometer and on a treadmill with at least 30 min rest between the two and when their heart rate returned to the previous baseline. The order was determined from random number tables. The cycle ergometer work rate was increased by 25 watts every minute to symptom limitation. The patients were all exercised on a cycle ergometer with increasing work rate of 15 watts every minute to symptom limitation. Data were recorded and stored on a breath by breath basis.

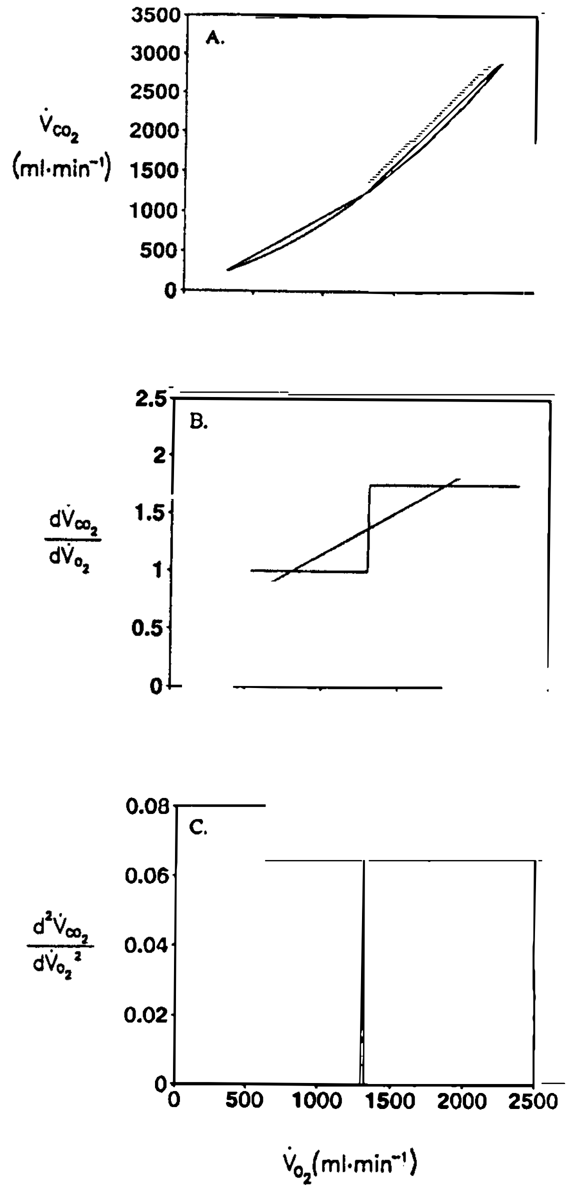

Figure 1. Panel A shows the functions used to generate data with and without GET. The ordinate is $\dot{V}CO_2$ and the abscissa is $\dot{V}O_2$ (ml/min). The two linear relations show a sudden change in the first derivative at the breakpoint. The curvilinear relation has no breakpoint. Panel B shows the first derivative ($d\dot{V}CO_2/d\dot{V}O_2$). The two linear relations show a sudden change in the first derivative at the breakpoint. The first derivative of the quadratic relation shows no abrupt alterations. Panel C shows the second derivative ($d^2\dot{V}CO_2/d\dot{V}O_2^2$). The second derivative of the two linear relations is zero except for the spike at the breakpoint. The second derivative of the quadratic relation shows no spike and remains zero.

Calculation of Results

Original V slope method (OVS). The original V slope method was im-

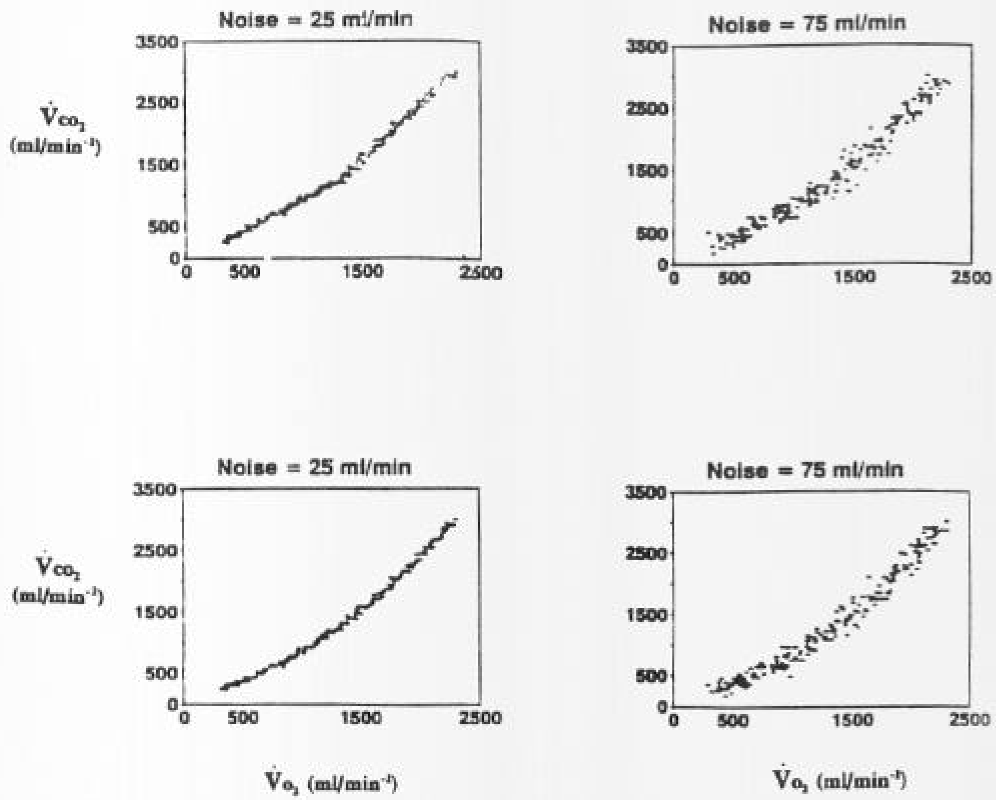

Figure 2. Four examples of simulated data sets with the highest slope ratio of 1.75. The noise level was 25 ml/min for the left hand panels and 75 ml/min for the right hand panels which represents the extremes of noise levels that we encountered in patients. Upper panels show data sets with a breakpoint and lower panels are the equivalent data sets without a breakpoint.

TABLE 1

CLINICAL DETAILS OF 15 SUBJECTS*

Subjects	Diagnosis	Age (yr)	FEV_1 (L)	FVC (L)	$\dot{V}O_{2max}$ (L/min)	HR_{max} (beats/min)	$\dot{V}E_{max}$ (L/min)
1 M	Normal	33	3.52 (83)	4.03 (79)	2.25 (72)	175 (93)	100 (81)
2 M	Normal	31	3.51 (80)	4.33 (83)	2.08 (65)	187 (99)	135 (110)
3 F	Normal	28	3.24 (101)	3.98 (108)	1.60 (72)	181 (94)	65 (57)
4 M	Normal	39	4.37 (103)	5.42 (99)	3.45 (117)	144 (78)	105 (69)
5 M	Normal	44	4.53 (108)	5.32 (102)	2.27 (81)	185 (102)	66 (42)
6 M	Unexplained dyspnea	70	2.81 (84)	3.64 (83)	1.25 (64)	116 (70)	56 (57)
7 M	Unexplained dyspnea	44	2.31 (60)	2.75 (58)	2.76 (99)	176 (97)	104 (129)
8 M	Exercised-induced asthma	29	4.67 (103)	5.7 (105)	1.70 (52)	159 (83)	63 (39)
9 M	Exercised-induced asthma	37	2.23 (59)	2.69 (60)	1.89 (63)	162 (87)	61 (78)
10 F	Chronic pulmonary embolism	59	2.56 (109)	3.02 (109)	1.16 (66)	153 (89)	42 (47)
11 M	COPD	72	2.06 (75)	3.46 (97)	1.40 (74)	169 (104)	76 (105)
12 F	Interstitial lung disease	40	1.07 (29)	1.36 (42)	0.81 (40)	131 (71)	91 (83)
13 M	Sarcoidosis	28	3.62 (85)	4.24 (85)	2.52 (77)	172 (90)	111 (88)
14 F	Exercise-induced asthma	17	2.82 (94)	3.33 (105)	2.16 (92)	207 (104)	48 (49)
15 F	Exercise-induced asthma	66	1.98 (106)	2.48 (111)	0.77 (46)	121 (73)	38 (55)

Definition of abbreviations: FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity; $\dot{V}O_{2max}$ = maximum oxygen uptake; HR_{max} = maximum heart rate; $\dot{V}E_{max}$ = maximum minute ventilation; COPD = chronic obstructive pulmonary disease; F = female; M = male.

* Exercise data are from exercise; % of predicted is shown in parenthesis.

plemented in precisely the same manner described previously (3); details are provided in Appendix A. In brief, it consists of two stages: data preparation and the breakpoint calculation. First, the data are smoothed with a moving (boxcar) average and the fluctuations in the breath to breath $\dot{V}CO_2$ values are corrected. Data are obtained during the first minute of exercise and above the respiratory compensation point. After data preparation, the breakpoint is calculated by using two adjoining linear regression relations until a best fit is found as judged by a modified least squares criterion. The criteria for detection of a breakpoint are given in the appendix.

Modified V slope method (MVS). The modified V slope method is a graphical procedure. For the simulated data, we used a group of 16 members of the Pulmonary and Critical Care Division at the State University of New York at Buffalo. For the simulated data the raters were given 18 graphs in random order with slopes of 1.1, 1.25, and 1.75 with noise levels of 5, 25, and 125 ml/min. Randomization was conducted by drawing cards from a box. For patient data, we used two physicians in our pulmonary fellowship program with experience in reporting clinical exercise tests and research. The raters were shown a plot of $\dot{V}CO_2$ and $\dot{V}O_2$ with equal axes and a line of identity. They were provided with a transparent ruler and asked to mark the point at which the data points were consistently different from a line parallel with the line of identity.

Nonparametric regression method (NPM). Nonparametric regression was performed using generalized additive models. The theoretical and practical implementation of this approach are given in reference 15; details specific for this study are provided in Appendix B. No data preparation is required. The regression relation is obtained without presupposing any particular mathematical form by using one or more adjoining cubic splines which are special types of cubic polynomials. One degree of freedom results in a linear relation, higher numbers will result in a more wiggly fit.

The best fit is one that has the least variance so that the regression is not altered much by individual data points, while exhibiting the least bias so that the data are scattered evenly around the regression line. This tradeoff between variance and bias is optimized by minimizing the generalized cross validation. The generalized cross validation is a quantity that is similar to the Akaike information criterion and Mallow's criterion for prediction (Cp) of standard multiple linear regression (15). The criteria for detection of a breakpoint are the same as those used for the original breakpoint.

This approach differs from the V slope method in several ways. First, the method is free to select the most appropriate fit; it is not forced to fit two linear relations. Second, nonparametric regression does not assume that the variance is uniform through the entire range (homoscedasticity). Unlike standard linear regression, the nonparametric regression is much less affected by aberrant points, particularly at the extremes of the range, which obviates the need for routine data preparation.

Statistical Methods

Comparisons of the detection accuracy, sensitivity, specificity and precision were evaluated with the Mantel-Haenszel test. A p value > 0.05 was considered not significant (NS). Numerical accuracy was compared using 95% confidence limits (17). The agreement on whether or not a threshold is present among raters performing MVS was assessed by kappa statistics (18). The number of correct responses between groups of raters was compared with Kruksal-Wallis one-way analysis of variance.

RESULTS

Simulated Data

Figure 3 shows the detection accuracy (correctness as to whether or not a threshold exists) of the three methods according to the amount of noise in the data and the slope ratio. NPM was significantly more accurate than OVS at noise levels of 5, 25, and 75

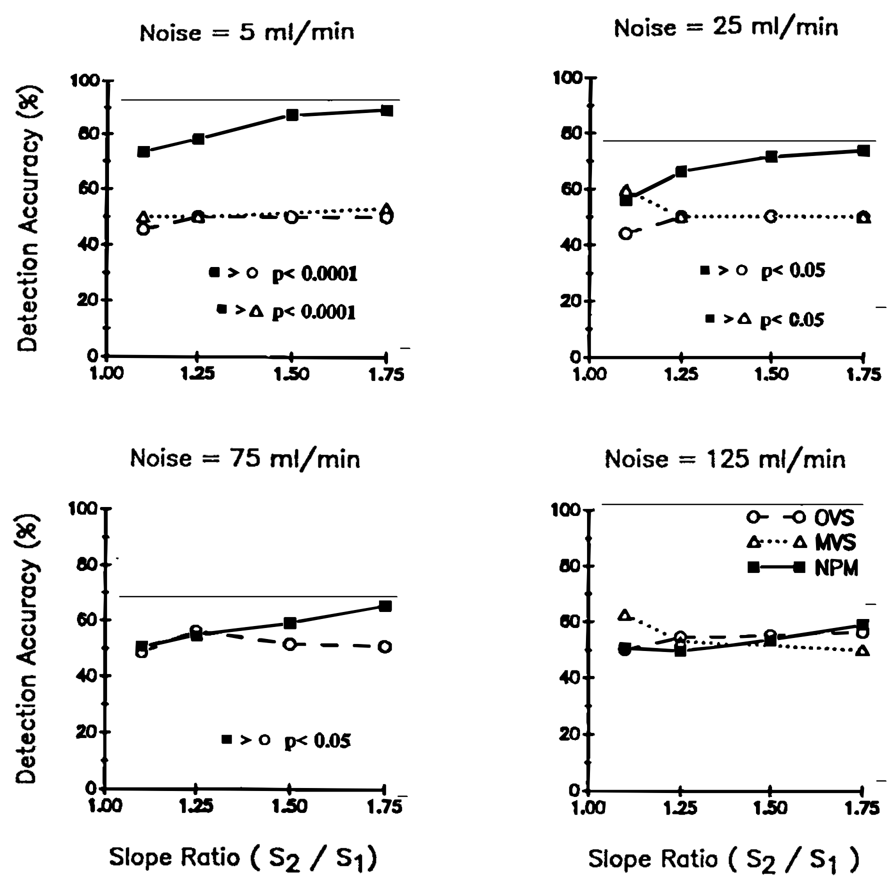

Figure 3. Detection accuracy of the three methods used to determine the gas exchange threshold. NPM had a significantly higher detection accuracy compared with OVS and MVS ($p < 0.05$) except at a noise level of 125 ml/min.

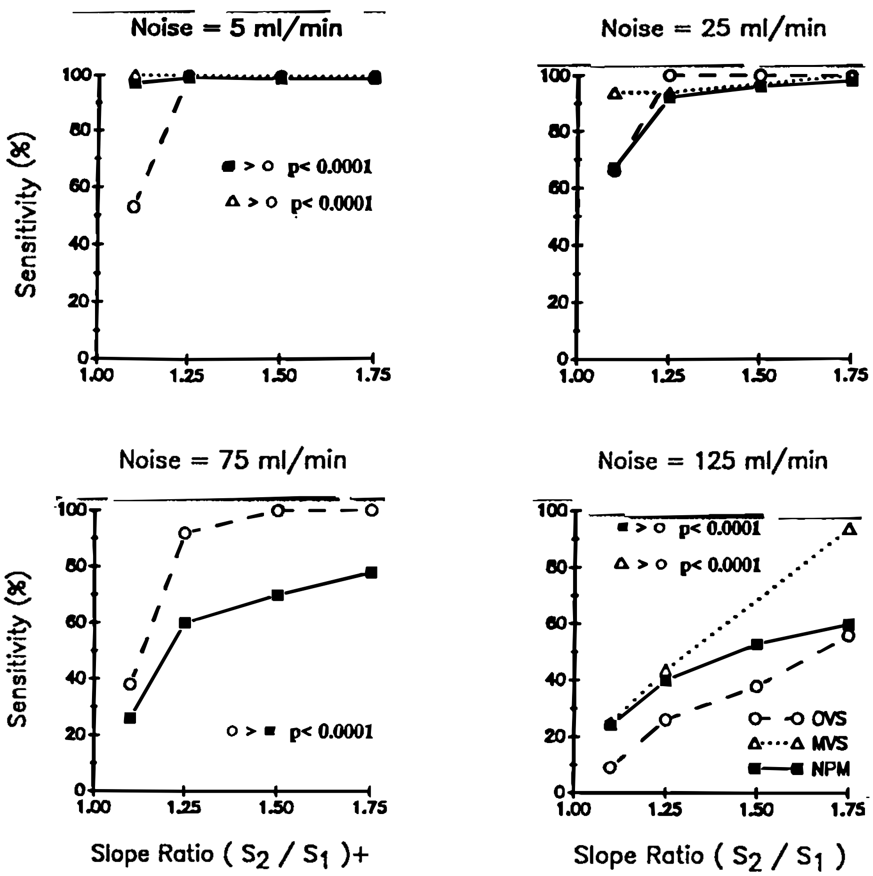

Figure 4. Sensitivity of the three methods used in the determination of GET. There were no significant differences between NPM and MVS. Both methods were more sensitive than OVS at noise levels of 5 and 125 ml/min ($p < 0.001$).

ml/min ($p < 0.0001$, $p < 0.0001$, and $p < 0.05$ respectively) and more accurate than MVS at noise levels of 5 and 25 ml/min ($p < 0.0001$ and $p < 0.05$ respectively). No measurements were made with MVS at 75 ml/min. The detection accuracy of NPM ranged between 50 and 89% depending on the amount of noise and slope ratio. There were no significant differences in detection accuracy between OVS and MVS. Detection accuracy never exceeded 63% for these two methods. All three methods had low detection accuracy and were not significantly different from each other at a noise level of 125 ml/min.

The sensitivity (true positive rate) of each method is shown in Figure 4. There were no significant differences between NPM and MVS at all compared noise levels. Both these methods were more sensitive than OVS at noise levels of 5 and 125 ml/min ($p < 0.001$). The difference at a noise level of 5 ml/min is solely due to a decreased sensitivity of OVS at a slope ratio of 1.1. OVS was more sensitive than NPM at a noise level of 75 ml/min ($p < 0.0001$). Figure 5 compares the specificity (true negative rate) of the three methods. Both OVS and MVS suffered from low specificity. NPM had a higher specificity than both OVS and MVS except at the highest level of noise (all $p < 0.0001$ except at a noise level of 125 ml/min).

Numerical accuracy of the three methods is compared in Figure 6. Accuracy is recognized by the true breakpoint being within the 95% confidence limits of the estimate, shown in the figure as error bars. Both NPM and OVS have similar degrees of numerical accuracy and are superior to MVS in this respect. The accuracy of the MVS estimate improved with increasing slope ratio. The mean differences from the true breakpoint for each method according to increasing levels of noise were: OVS −2, −28, −128, and −13 ml/min; MVS 128, 192, and 348 ml/min; NPM 1, −9, −30, and −90 ml/min. MVS overestimated the true breakpoint at all levels of noise. There were no significant differences among the three methods with regard to precision which is expressed as the normalized standard error.

Agreement Among Raters

The overall kappa value among all raters performing MVS as to the presence of a threshold was 0.53 indicating only a modest amount of agreement (19). To analyze whether experience makes a difference in the detection of a threshold, raters were divided into three groups: pulmonary fellows (n = 6), attending physicians who report pulmonary function tests on a regular basis (n = 5), and attending physicians who do not report pulmonary function tests (n = 5). The kappa for each group was 0.76, 0.47, and 0.41 respectively. The percentage of correct responses for each of these groups was 49% ± 1 (SE), 52% ± 1 (SE), and 54% ± 2.5 (SE). These scores were not significantly different.

Human Data

All three methods detected and gave comparable estimates of GET on 10 exercise tests performed by normal subjects as shown in Table 2. Similar results were obtained in 7 of the 10 patients (Table 2). OVS invariably detected a threshold on all patients, whereas NPM and MVS did not detect a threshold in two patients. Estimates of GET by OVS were lower ($p < 0.05$ after Bonferroni cor-

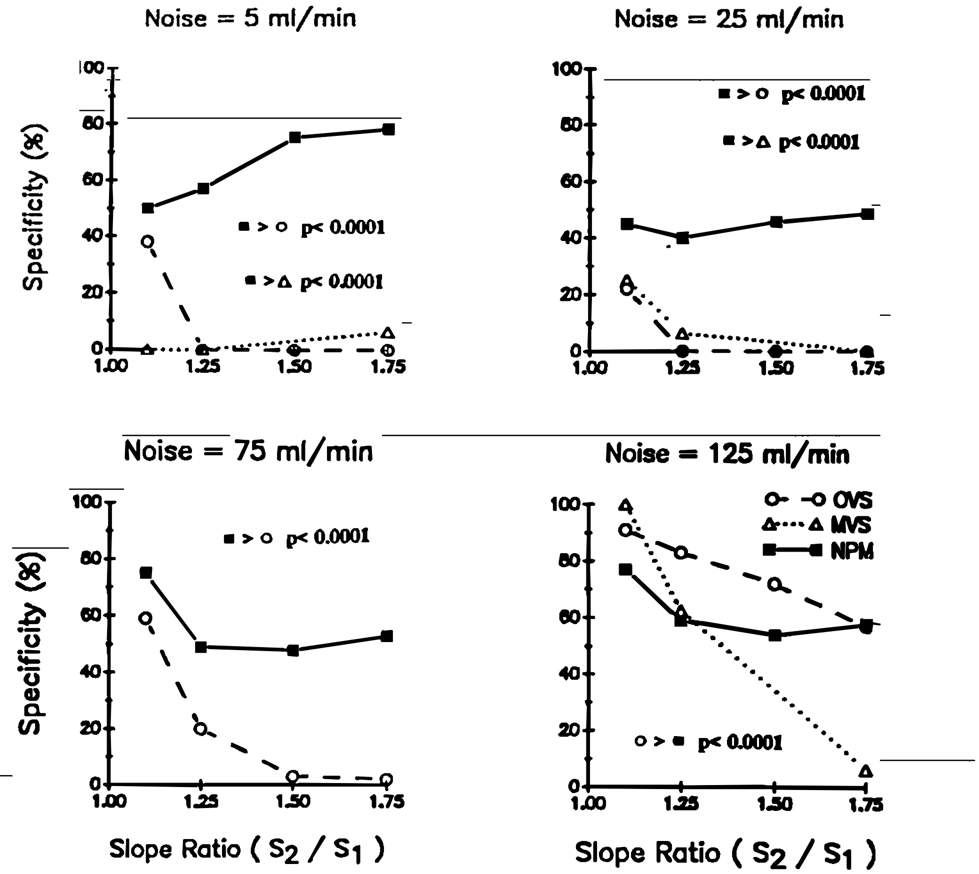

Figure 5. Specificity of the three methods used in the determination of GET. NPM had a higher specificity than both OVS and MVS except at the highest level of noise ($p < 0.0001$). Both OVS and MVS suffered from low specificity.

rection) than for MVS (mean difference −368 ml/min, 95% confidence limits −539 to −197 ml/min) and NPM (mean difference −336 ml/min, 95% confidence limits −559 to −114 ml/min). This difference was primarily due to elimination of data above the respiratory compensation point. When the procedure is eliminated from OVS, the mean difference between estimates by OVS and MVS was 58 ml/min (95% confidence limits −118 to 234 ml/min, n = 18), 89 ml/min (95% confidence limits −17 to 196 ml/min, n = 18) by OVS and NPM, and 22 ml/min (95% confidence limits −149 to 192 ml/min, n = 17) for MVS and NPM. None of these differences were significantly different from zero.

TABLE 2

COMPARISON OF GET ESTIMATED BY OVS, MVS, and NPM

Subjects	OVS (*ml/min*)	MVS (*ml/min*)	NPM (*ml/min*)
1 C	886	1,178	1,793
T	1,244	1,658	2,004
2 C	914	1,079	1,720
T	1,171	1,652	1,797
3 C	546	801	698
T	789	1,164	806
4 C	1,190	2,518	2,623
T	2,027	3,196	2,900
5 C	978	1,427	1,037
	1,420	1,855	1,859
6 C	847	ND	716
7 C	1,276	1,469	1,546
8 C	765	905	684
9 C	965	1,375	881
10 C	885	852	954
11 C	903	905	1,068
12 C	642	ND	ND
13 C	1,104	1,348	1,096
14 C	969	1,208	759
15 C	548	613	ND

Definition of abbreviations: GET = gas exchange threshold; OVS = original V-slope method; MVS = modified V-slope method; NPM = nonparametric method; C = cycle ergometry; T = treadmill exercise; ND = GET was not detected.

Figure 7 shows examples of detection of anaerobic threshold using NPM. For Subject 5, GET is defined easily at 1,037 ml/min and it is associated with an obvious peak of the second derivative. For Subject 11, GET is difficult to define but is present at 1,068 ml/min. The peak of the second derivative is somewhat blunted compared with that in Subject 5. For Subject 12, GET cannot be defined by NPM because no spike can be identified.

DISCUSSION

This study compared the test characteristics of three methods used in the detection of GET. OVS is an objective method that uses two linear regressions, MVS is a visual method, and NPM is a new objective method that we developed which utilizes nonparametric rather than linear regression. Comparisons were made on simulated data where a true threshold was known with absolute certainty because there is no gold standard at present to serve as the reference test. NPM was found to be significantly more accurate than OVS and MVS for detecting GET (detection accuracy) except with high levels of noise. The detection accuracy for OVS and MVS never exceeded 63% because these two methods suffered from a low specificity compared with NPM. As a result of this improved specificity, the detection accuracy of NPM ranged

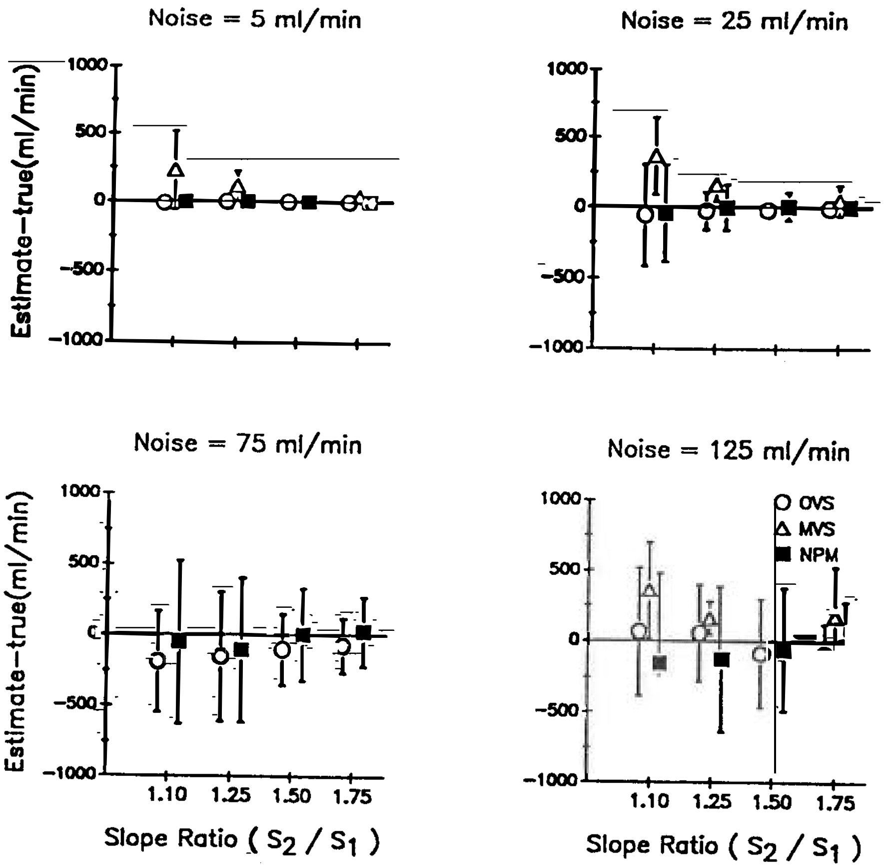

Figure 6. Numerical accuracy of the three methods used in the determination of GET. Mean values are shown with their 95% confidence limits represented as error bars. Both NPM and OVS had similar degrees of numerical accuracy. Both were superior to MVS.

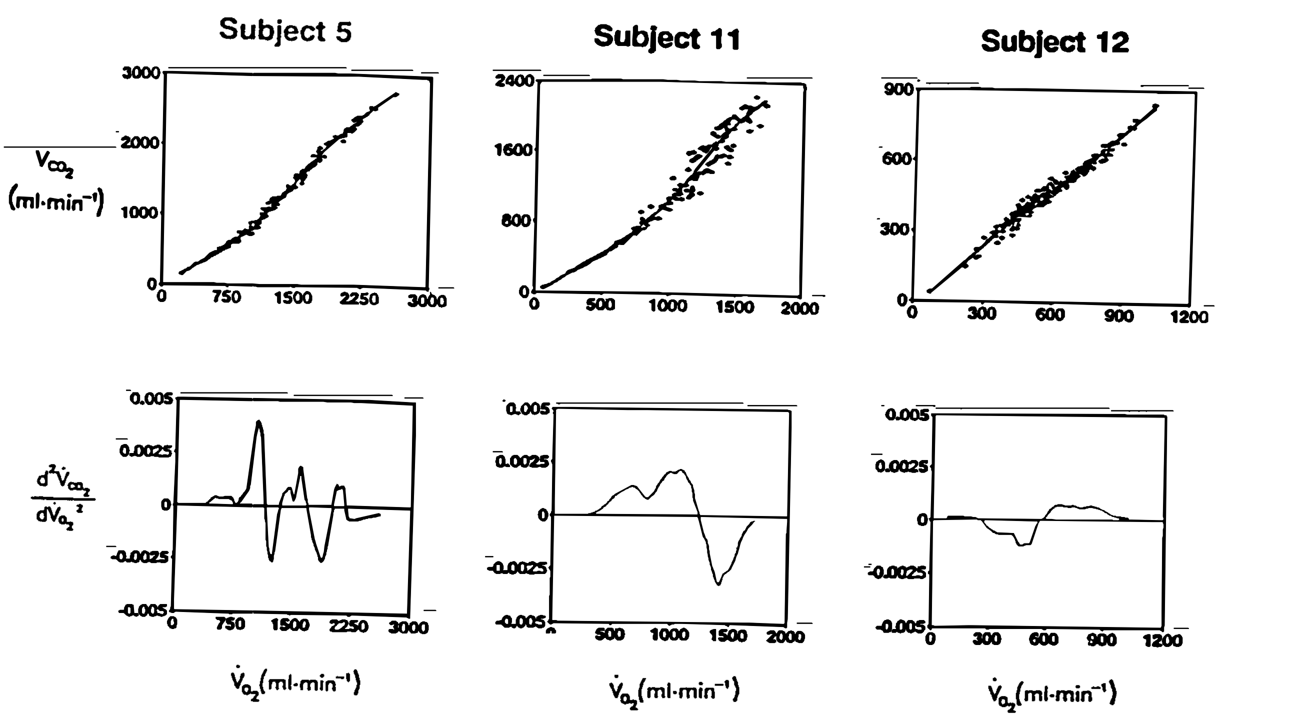

Figure 7. Examples of $\dot{V}CO_2$ - $\dot{V}O_2$ relation (*upper panels*) and of its second derivative ($d^2\dot{V}CO_2/d\dot{V}O_2^2$) plotted against $\dot{V}O_2$ (*lower panels*) obtained with the nonparametric method during cyclometer exercise in three subjects referenced in Tables 1 and 2.

between 50 and 89% depending on the amount of noise and abruptness of the threshold.

The sensitivity of NPM was as accurate as OVS with the exception of a noise level of 75 ml/min. The simulated data with a breakpoint were generated from two linear equations with a uniform variance around these relations throughout the range of the data. Therefore, the OVS which uses two adjoining linear regression relations is ideally suited to detect breakpoints from these data sets. Therefore, it is surprising to us that the sensitivity of NPM was comparable to that of OVS. As Figure 7 illustrates, data obtained from patients in a clinical setting do not follow this ideal pattern. The variance in the data from the fitted line may not be uniform over the range of measurement as indeed is the case for Subject 11 (Figure 7). Therefore, the use of two adjoining linear relations may not be applicable to real data.

The low specificity of OVS is attributed to the fact that the two linear regression relations are well suited to fit curvilinear data. As a result, it invariably detects a threshold. However, estimates of the threshold obtained by OVS are as close to the true value as NPM. MVS also had low specificity. This result confirms Belman and coworkers (14) who suggested that the specificity of MVS was low at 77% although their 95% confidence limits were wide (55–92%). In addition, we found that MVS was inferior to the other two methods in terms of numerical accuracy though it had comparable degrees of precision. The normalized standard error (standard error of the mean divided by the true mean) of MVS ranged from 4 to 11% with low levels of noise. MVS overestimated the true threshold at all levels of noise.

Agreement on the presence of a threshold was only modest among all raters who performed MVS. Furthermore, visual detection of a threshold appears not to be dependent on experience because pulmonary fellows scored as well as attending physicians, whether or not they reported pulmonary function tests regularly.

NPM has the following advantages in the detection of GET. First, the procedure does not routinely require special data preparation or observer intervention. It is objective and therefore an efficient approach for clinical studies and avoids the introduction of potential bias in research studies. Second, the accuracy of detecting a threshold is improved because of higher specificity compared with the two currently used methods. Although these improvements were lost at a noise level of 125 ml/min, analysis of the noise level in our human data revealed that they ranged from approximately 25 to 75 ml/min. Third, the validity of the evaluation of GET in each subject can be quantified and clearly displayed graphically with the second derivative plot.

A disadvantage of NPM is that it is dependent on numerical differentiation for which there is no formal method for testing its accuracy. Like OVS, the assumption used in NPM is that the most abrupt change that meets specific criteria is due to the gas exchange threshold. The methodology presented here only addresses the identification of GET from the $\dot{V}CO_2$-$\dot{V}O_2$ plot. On occasions, additional information may be required such as ventilatory equivalents for O_2 and CO_2 and end-tidal O_2 and CO_2 partial pressures.

This study highlights the difficulty of all methods that attempt to define a breakpoint from discrete data. The concept of a threshold (breakpoint) is defined clearly for continuous data. A breakpoint can be defined as the locus where the second derivative cannot be defined (a singularity). The transfer of this definition to discrete data is problematic. The data from standard clinical exercise tests are by nature discrete because they are limited to breath by breath collections or discrete blood samples. A breakpoint can be surmised from discrete data by an abrupt change in the second derivative. However, with noise in the data, the detection of a breakpoint may become blurred, hence the definition of an abrupt change becomes necessary. The criteria for an abrupt change that we used for NPM are equivalent to those of Beaver and associates (3). Other criteria might be developed. It should be emphasized that none of our normal subjects or patients were experienced in clinical exercise testing. Therefore, the data obtained from these subjects are representative of usual operating conditions in pulmonary function laboratories.

In conclusion, we have shown with the use of simulated data that the detection accuracy of both OVS and MVS suffers because of low specificity. A new method using NPM had superior detection accuracy compared with the previous methods because of improved specificity. The results of this study provide further evidence to support the existence of a breakpoint occurring in gas exchange data during exercise but it does not test the concept of an anaerobic threshold. Problems with reliability of GET are at least in part due to variation in its detection. If GET is to be widely used clinically, an objective and accurate method for its detection is necessary. We believe that NPM offers substantial improvements in this regard compared with existing methods. Problems in its detection are readily apparent from plots of the second derivative against $\dot{V}O_2$. However, its validity needs to be confirmed with prospective studies that relate its estimates to the lactate or lactic acidosis thresholds.

APPENDIX

A. Original V Slope Method (OVS)

The first stage is the data preparation; it is only required for the human data. The breath by breath $\dot{V}CO_2$ and $\dot{V}O_2$ data were smoothed with a 9 s moving average. The breath by breath end-tidal PCO_2 data plotted against time were fitted with a cubic polynomial regression. The deviation of each end-tidal PCO_2 from its predicted value from the cubic expression was used to adjust the $\dot{V}CO_2$ values as described previously with an iterative procedure. A golden section search method was used for this purpose (16). Two linear regressions are fitted to the breath by breath minute ventilation and $\dot{V}CO_2$ to determine the respiratory compensation point (RC). RC is determined as the intersection between the two linear segments provided the slope of the second segment is 15% greater than the slope of the first segment. Data were eliminated from the first minute of exercise and at a $\dot{V}CO_2$ above the RC. When no RC was detected or it was clearly erroneous, only data obtained during the first minute of exercise were eliminated.

The second stage is to fit two linear regression equations to the processed data of $\dot{V}CO_2$ and $\dot{V}O_2$ to determine the tentative value for GET at the point of their intersection. The tentative GET was used as an initial guess for an iterative procedure: the downhill simplex method (16). The two regression relations were varied to maximize the ratio of the distance from the intersection point to the linear relation of the entire data set to the mean residual sum of squares. The slope of the lower segment (S1) was greater than 0.6; therefore it was not necessary to eliminate data after the first minute of exercise (3). The criterion for a threshold is that the slope of the upper segment must be 1.1 times S1 or greater.

B. Nonparametric Regression Method (NPM)

To determine the best fit, the number of degrees of freedom that are assigned to the regression is initially one, and progressively increased in unit steps until the generalized cross validation no longer decreases. The detection and numerical identification of GET are based on the second derivative of the fitted regression line plotted against $\dot{V}O_2$. The first and second derivatives are calculated with Lagrangian second order polynomials (16) from the following expressions where n indicates the breath number and N is the total number of breaths collected during the exercise period.

$$d\dot{V}CO_2/d\dot{V}O_2(n) = f[\dot{V}CO_2(n-1), \dot{V}O_2(n-1); \dot{V}CO_2(n), \dot{V}O_2(n); \dot{V}CO_2(n+1), \dot{V}O_2(n+1)]$$

$$d^2\dot{V}CO_2/d\dot{V}O_2{}^2(n) = f[d\dot{V}CO_2/d\dot{V}O_2(n-1), \dot{V}O_2(n-1); d\dot{V}CO_2/d\dot{V}O_2(n), \dot{V}O_2(n); \dot{V}CO_2/d\dot{V}O_2(n+1), \dot{V}O_2(n+1)]$$

A linear relation between $\dot{V}CO_2$ and $\dot{V}O_2$ would result in a zero second derivative. A breakpoint would result in an abrupt increase in the second derivative. Therefore a breakpoint was identified if the second derivative became positive and then became nonpositive at a higher $\dot{V}O_2$. The $\dot{V}O_2$ at which the second derivative is maximal within that interval is the value of the potential breakpoint.

For GET detection, the data are discrete and subject to noise because they are collected on a breath by breath basis. Therefore, a sharp spike does not occur, and criteria are needed to define whether or not a change in the first derivative is considered abrupt. To determine if the breakpoint represented a significant threshold, the criteria were equivalent to those used by Beaver and coworkers. (3). In terms of nonparametric regression, these criteria are equivalent to the value of the first derivative being greater than 0.6 at the $\dot{V}O_2$ when the second derivative becomes positive (D_1); and the first derivative at the $\dot{V}O_2$ when the second derivative becomes nonpositive must be greater than 1.1 times D_1. If more than one breakpoint is identified, GET was selected as the one with the highest value of the second derivative.

References

1. Wasserman, K., and M. B. McIlroy. 1964. Detecting the threshold of anaerobic metabolism in cardiac patients during exercise. *Am. J. Cardiol.* 14:844–852.
2. Patessio, A., R. Casaburi, M. Carone, L. Appendini, C. F. Donner, and K. Wasserman. 1993. Comparison of gas exchange, lactate and lactic acidosis thresholds in patients with chronic obstructive pulmonary disease. *Am. Rev. Respir. Dis.* 148:622–626.
3. Beaver, W. L., K. Wasserman, and B. J. Whipp. 1986. A new method for detecting anaerobic threshold by gas exchange. *J. Appl. Physiol.* 60:2020–2027.
4. Sue, D. Y., K. Wasserman, R. B. Moricca, and R. Casaburi. 1988. Metabolic acidosis during exercise in patients with chronic obstructive pulmonary disease. *Chest* 94:931–938.
5. Koike, A., M. Hiroe, K. Taniguchi, and F. Marumo. 1993. Respiratory control during exercise in patients with cardiovascular disease. *Am. Rev. Respir. Dis.* 147:425–429.
6. Kinker, J. R. 1993. Lactate breakpoint during slowly increasing work rates and irregular breathing. *Am. Rev. Respir. Dis.* 147(3):761–763.
7. Older, P., R. Smith, P. Courtney, and R. Hone. 1993. Preoperative evaluation of cardiac failure and ischemia in elderly patients by cardiopulmonary exercise testing. *Chest* 104:701–704.
8. Wasserman, K. 1984. The anaerobic threshold measurement to evaluate exercise performance. *Am. Rev. Respir. Dis.* 129(suppl):S35–S40.
9. Itoh, H., K. Taniguchi, A. Koike, and M. Doi. 1990. Evaluation of severity of heart failure using ventilatory gas analysis. *Circulation* 81(suppl II):II-31–II-37.
10. Casaburi, R., A. Patessio, F. Ioli, S. Zanaboni, C. F. Donner, and K. Wasserman. 1991. Reductions in exercise lactic acidosis and ventilation as a result of exercise training in patients with obstructive lung disease. *Am. Rev. Respir. Dis.* 143:9–18.
11. Hughson, R., K. Weisiger, and G. Swanson. 1987. Blood lactate concentration increases as a continuous function in progressive exercise. *J. Appl. Physiol.* 62:1975–1981.
12. Yeh, M., R. Gardner, T. Adams, F. Yanowitz, and R. Crapo. 1983. "Anaerobic threshold": problems of determination and validation. *J. Appl. Physiol.* 55:1178–1186.
13. Gladden, B. L., J. W. Yates, R. W. Stremel, and B. A. Stamford. 1985. Gas exchange and lactate anaerobic thresholds: inter and intraevaluation agreement. *J. Appl. Physiol.* 58:2082–2089.
14. Belman, M., L. Epstein, D. Doornbos, J. Elashoff, S. Koerner, and Z. Mohsenifar. 1992. Non-invasive determinations of the anaerobic threshold reliability and validity in patients with COPD. *Chest* 102:1028–1034.
15. Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. New York: Chapman and Hall, 136–173.
16. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1986. Numerical recipes. Cambridge: Cambridge University Press, 498–546.
17. Bland, J. M., and D. G. Altman. 1986. Statistical methods for assessing agreement between two methods of clinical assessment. *Lancet* 1:307–310.
18. Davies, M., and J. Fleiss. 1982. Measuring agreement for multinomial data. *Biometrics* 38:1047–1051.
19. Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.