Point of View

Optimizing the Exercise Test for Pharmacological Investigations

Jonathan Myers, MA, and Victor F. Froelicher, MD

Exercise trials in cardiology are often hindered by inconsistent approaches to exercise testing. These inconsistencies include the choice of exercise protocol, exercise end points, points of analysis, and absence or misuse of gas exchange data. Gas exchange techniques greatly enhance the accuracy with which cardiopulmonary function is assessed by exercise. Commonly used protocols are not always appropriate for all patients or all studies. Both cardiovascular disease and the exercise protocol can have an important impact on the relation between changes in work rate and oxygen uptake. Ramp protocols appear to offer the greatest promise for assessing cardiopulmonary function. Analyzing hemodynamic and gas exchange responses at several points submaximally, in addition to those at peak exercise, can add important information concerning the efficacy of a drug. A great deal of confusion continues to hinder the application of the gas exchange anaerobic threshold, and many of the commonly used testing end points are not reliable. (Circulation 1990;82:1839–1846)

Exercise testing is routinely used clinically to assess the presence and severity of cardiovascular disease. Maximal exercise capacity, which is optimally quantified by maximal ventilatory oxygen uptake, defines the upper limit of cardiorespiratory function. Exercise test results are frequently the sole criterion used to evaluate the efficacy of a given pharmacological intervention. Knowledge concerning the physiological responses to acute and chronic exercise has undergone a rather remarkable evolution in recent years. Nevertheless, results of pharmacological studies are frequently confounded by methodological differences including exercise protocols, exercise end points, absence or misuse of gas exchange data, and study design. Because pharmaceutical companies frequently submit the results of multicenter exercise testing trials for consideration by the Food and Drug Administration, consistent and sound principles of exercise testing, and study design are paramount in importance. The purpose of this communication is to review these issues with the aim of optimizing the exercise test for pharmacological investigations.

Selection of Patients: Screening, Reproducibility, Criteria

The quality of studies that use exercise to objectively evaluate the efficacy of a drug, particularly those that involve angina patients, depends heavily on careful screening of patients before qualification. Because angina can vary greatly in some patients, it is important to select stable patients whose symptoms are known to occur reproducibly in response to exercise. Previous investigations have observed a tendency for patients to increase exercise time in a standard protocol with serial testing on the same day and on different days, even with several weeks between tests. Changes in treadmill time with serial testing have even been observed without changes in maximal heart rate or double product. Sullivan et al, however, did not observe any differences in maximal oxygen uptake among three treadmill tests performed in 1 week although there was a significant increase in treadmill time.

A common method of assuring a reproducible response to exercise is to have the patient perform at least two exercise tests on separate days, at the same time of day. A test is considered “reproducible” if peak oxygen uptake is within ±10% on both days. In patients without prior treadmill experience or without experience with gas exchange equipment, or without both, it is preferable to use the first test for habituation purposes, and judge the patients suitability from subsequent tests. Elborn et al performed three consecutive treadmill tests on separate days in patients with heart failure, and reported that the first test underestimated exercise time by approximately 20%. Pinsky and associates performed repeated treadmill tests among patients with heart failure until test duration on three consecutive tests varied by less than 60 seconds. This stability criteria was achieved

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.
From the Cardiology Section, Long Beach Veterans Administration Medical Center, Long Beach, California.
Supported in part by Burdick Corporation.
Address for reprints: Victor F. Froelicher, MD, Chief, Cardiology Department (111C), Long Beach Veterans Administration Medical Center, 5901 East Seventh Street, Long Beach, CA 90822.
within three tests on only nine of 30 patients, whereas 13 patients required four or five tests and eight patients required more than six tests. The stability of exercise time among patients with angina has not been satisfactorily reported.

Although this issue has not been resolved adequately, the inclusion of a reproducibility criterion is nevertheless essential as it increases confidence in the study results and reduces the chances of making a type I error (i.e., concluding a drug has had an effect when it has not). Although we recommend all serial testing be performed on separate days, results have been reported in which exercise testing was performed 15 minutes to 10 hours apart on the same day.1,7,8 The issue of diurnal variation is addressed herein. Screening exercise tests also provide an opportunity to evaluate other considerations concerning the study, that is, patient stability and any untoward responses such as rhythm disturbances, blood pressure abnormalities, or other reasons for stopping exercise. Angina or heart failure patients who are limited by other problems such as claudication or lung disease are not suitable for drug studies.

For angina studies, it is preferable for patients to exhibit both subjective (angina) and objective (ST segment changes) markers of ischemia during exercise. During screening, all patients should exhibit ST segment changes consistent with a positive exercise test; a minimum of 1.0 mm exercise-induced ST segment depression below the resting PR segment, which is horizontal or downsloping, should be included in the qualification criteria. Changes isolated to the inferior leads (II and aVF) are often false positives, and patients with left bundle branch block should be excluded.

Mode of Exercise

Studies comparing treadmill and bicycle exercise have reported maximal oxygen uptake to be generally 10–15% higher on the treadmill (range, 6–25%).9–11 Heart rate has been reported to be from 5% to 20% higher on the treadmill. Not surprisingly, ST segment changes have been reported to be somewhat more frequent during treadmill testing.12 Thus, for studies in which the efficacy parameters include the functional limits of the patient and objective signs of ischemia, the treadmill is preferable.

Respiratory Gas Exchange Techniques

In drug studies, maximal oxygen uptake is usually estimated from treadmill or bicycle time or work. This practice, which is typical of most clinical exercise testing, is fraught with errors, which have been attributed to differences in protocols, efficiency, fitness, and the presence of disease.4,13–16 Although maximal oxygen uptake and treadmill time have been closely related (with correlation coefficients reported to range between 0.80 and 0.90),4,16,17 a wide scatter around these regression lines has been observed.14,17 Froelicher et al17 demonstrated that treadmill time can change significantly without a change in oxygen uptake. Among over 1,000 asymptomatic men, for any given treadmill time on the Balke protocol, the 95% confidence limits for oxygen uptake ranged nearly 20 ml/kg/min. Other investigators13–15,18 have since confirmed the shortcomings associated with predicting oxygen uptake from treadmill time. Sullivan and McKirnan13 and Roberts and associates14 reported that among patients with coronary artery disease, oxygen uptake was 13% and one MET lower than normal patients, respectively, for the same treadmill work at higher levels of exercise. Because a favorable effect of a given medication on exercise capacity often ranges 10–20%,19–21 the potential inaccuracies associated with predicting ventilatory oxygen uptake from treadmill time become obvious.

Sullivan and associates4 later reported that measured oxygen uptake was more reproducible than treadmill time in patients with angina who were participating in a drug study. Measured oxygen uptake had a higher reliability coefficient (r=0.88) than treadmill time (r=0.70) across three exercise tests on different days. The 90% confidence intervals for the reliability coefficients were higher for measured oxygen uptake (r=0.76–0.95) than for treadmill time (r=0.48–0.86). Thus, the use of gas exchange techniques, by providing a more accurate and reproducible measure of cardiopulmonary function, can greatly improve the quantification of the effect of a drug in clinical studies. Recent technological advances allow on-line assessment of oxygen uptake and other gas exchange parameters, which facilitate precision and convenience.

Exercise Protocol

Stuart and Ellestad in 198022 surveyed 1,375 exercise laboratories in North America and reported that of those performing treadmill testing, 65.5% use the Bruce protocol for routine clinical testing. This protocol uses uneven 2 to 3 MET increments in work every 3 minutes. Investigators have since recommended protocols with smaller and more equal increments.1,11,23 Redwood and associates performed serial testing in patients with angina and reported that work rate increments that were too rapid resulted in a reduced exercise capacity, and could not be reliably used for studying the effects of therapy. When excessive work rates were used, the reduction in myocardial oxygen demand due to nitroglycerin was minor, suggesting that protocols placing heavy and abrupt demands on the patient may mask a potential salutary effect of an intervention. These investigators recommended that the protocol be individualized for each patient to elicit angina within 3–6 minutes. Smokler and associates24 reported that among 40 pairs of treadmill tests conducted within a 6-month period, tests that were less than 10 minutes in duration showed a much greater percentage of variation than those that were greater than 10 minutes in duration. Buchfuhrer et al11 performed repeated maximal exercise testing in five normal subjects while varying the work rate increment. Max-
imal oxygen uptake varied with the increment in work; the highest values were observed when intermediate increments were used. These investigators suggested that a test with the work increments individualized to yield a duration of approximately 10 minutes was optimal for assessing cardiopulmonary function.

A rather novel protocol in which the work rate increases progressively but has no "increments" per se is the ramp test. The ramp protocol uses a consistent and continuous increase in external work. Whipp et al \(^{25}\) demonstrated a linear relation between oxygen uptake and work rate by using a ramp bicycle protocol, an observation not characteristic of the staged protocols such as the Bruce test.\(^{13,18}\) The potential advantages of the ramp test are attractive. Although few laboratories performing drug studies to our knowledge have used the test, it appears to offer several advantages for cardiopulmonary assessment. The recommendations concerning smaller work rate increments, individualizing the test, and optimizing test duration would be facilitated by a ramp test.

To investigate this, our laboratory recently compared ramp treadmill and bicycle tests to protocols more commonly used clinically.\(^{26,27}\) Ten patients with chronic heart failure, 10 with coronary artery disease who were limited by angina during exercise, 10 with coronary artery disease who were asymptomatic during exercise, and 10 age-matched normal subjects performed three bicycle tests (25 W/2-min stage, 50 W/2-min stage, and ramp) and three treadmill tests (Bruce, Balke, and ramp). Ramp rates on the bicycle and treadmill were individualized to yield a test duration of approximately 10 minutes for each subject. Maximal oxygen uptake was significantly higher on the treadmill protocols versus the bicycle protocols collectively, confirming previous observations.\(^{9-11}\) Only minor differences in maximal oxygen uptake, however, were observed between the treadmill protocols. The relations between oxygen uptake and work rate (predicted oxygen uptake), defined as a slope, were highest for the ramp test and poorest for the tests with the largest increments in work (Bruce treadmill and 50 W/stage bicycle). Additionally, the SEE (oxygen uptake, ml/kg/min) for oxygen uptake versus workload or time was largest for the Bruce test and smallest for the ramp test. These observations suggest that 1) oxygen uptake is overestimated from tests with large increments in work, and 2) the variability in estimating oxygen uptake from work rate is markedly greater on these tests than for an individualized ramp treadmill test.

It is also interesting how oxygen uptake kinetics were influenced by the presence of disease. The oxygen uptake slopes were generally steeper (closer to unity) among normal subjects regardless of the protocol used. Patients with poor left ventricular function (chronic heart failure) overall had the poorest relation between oxygen uptake and work rate. This confirms the observations of Roberts et al \(^{14}\) who reported a reduced slope in oxygen uptake versus treadmill work as exercise progressed among patients with resting ejection fractions of less than 0.40. We observed a pronounced improvement, however, in the slope of oxygen uptake in these patients when using an individualized protocol. In fact, the response of patients with heart failure was similar to that of normal subjects when both groups performed ramp treadmill tests.

These observations emphasize the importance of measuring, rather than predicting, oxygen uptake when studying treatment effects. Several investigations have confirmed that substantial errors can result when using certain protocols or studying certain disease states, or both. It appears that the ramp protocol offers promise for optimizing cardiopulmonary assessment when evaluating pharmacological or other interventions. A growing number of investigators have discussed the importance of adapting the exercise test to the subject and purpose of the test.\(^{1,11,28}\) reducing the work increment appears to offer the best opportunity to observe a treatment effect.\(^1\) Because ramp rates can be adjusted as small or large as needed to suit a given patient or test purpose, and by definition increments are negligible, the ramp test appears ideally suited for pharmacological studies. More data are necessary, however, to confirm its usefulness in clinical trials.

One impediment to widespread application of the ramp is judging the rate of work for each patient so that exercise time is optimized at approximately 10 minutes,\(^{11,28}\) or a test duration chosen by the investigator. The specific activity scale outlined by Goldman et al \(^{29}\) and the Duke activity questionnaire\(^{30}\) offer potential for estimating functional capacity beforehand so that the ramp rate can be set appropriately for a given patient. Alternatively, when available, the results of a recent exercise test can be used.

**Laboratory Conditions**

Conditions in the laboratory that should be considered include temperature, ventilation, and time of day. Temperature changes are known to be one source of angina. Lassvik and Areskog\(^{31}\) reported a relatively minor 7% reduction in exercise time when patients with angina were tested at \(-10^\circ\) C versus \(-22^\circ\) C. Juneau et al\(^{32}\) studied eight "cold-sensitive" patients, and reported a 30% decrease in exercise time to 1.0 mm ST segment depression, and a 16% reduction in exercise time to the onset of angina at \(-8^\circ\) C versus \(20^\circ\) C. The laboratory should also be well ventilated. This is particularly important when collecting gas exchange data; these tests are often performed with several technical persons and a physician present. In a small or unventilated room, this can alter the concentration of inspired air. We frequently point a fan in the general direction of the patient's inspiratory valve to assure an accurate fraction of inspired oxygen. For a given patient participating in a clinical trial, testing must also be performed at the same time of day. Although the major concern is performing exercise testing consistently in
terms of the time medication was taken, the issue of diurnal variation has been repeatedly raised in normal subjects\textsuperscript{7,33} and in patients with angina.\textsuperscript{8,34,35} Garrard and Emmons\textsuperscript{7} performed morning and afternoon exercise testing in the same subjects on different days, and reported that afternoon testing was associated with consistently greater values for certain gas exchange variables. Handler and Sowton\textsuperscript{34} reported that markers of ischemia (ST changes, angina, or both) occurred reproducibly whether testing was performed in the afternoon or in the morning. Joy and Pollard,\textsuperscript{8} however, reported significant variation in ST segment and hemodynamic responses among patients with stable angina who performed exercise testing 8 hours apart on the same day.

**Points of Analysis**

In addition to data obtained at rest, a number of points during exercise, both maximal and submaximal, can yield important information concerning treatment effects. These points are listed in Table 1. A matched (placebo versus drug) submaximal workload is a particularly important analysis when studying, for example, oxygen uptake in patients with congestive heart failure or a \( \beta \)-blocker among patients with hypertension or angina. Under normal conditions, the oxygen requirements for a given submaximal level of work are approximately the same, even between patients or before and after exercise training. Patients with congestive heart failure, however, are known to have a reduced cardiac output for a given level of submaximal work. Because most medications have the primary goal of increasing contractility or reducing afterload, or both, in these patients, a comparison of the oxygen requirements at a matched submaximal workload can add significantly to the analysis. Likewise, the negative inotropic effects of \( \beta \)-blockade and their influence on the hemodynamic and gas exchange response to submaximal work should be included in studies using these drugs in patients with angina or hypertension.

<table>
<thead>
<tr>
<th>TABLE 1. Points of Comparison Between Placebo and Drug Phases in Clinical Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resting heart rate, blood pressure</td>
</tr>
<tr>
<td>Matched submaximal workload</td>
</tr>
<tr>
<td>Matched rate-pressure product</td>
</tr>
<tr>
<td>Onset of angina</td>
</tr>
<tr>
<td>Moderately severe (3 on 1-to-4 scale) angina</td>
</tr>
<tr>
<td>Gas exchange anaerobic threshold</td>
</tr>
<tr>
<td>Maximal exertion</td>
</tr>
</tbody>
</table>

In angina studies, we commonly match an arbitrary rate-pressure product, individualized for each patient, as late as possible during the exercise test. Rate-pressure product is the most accurate noninvasive index of myocardial oxygen demand,\textsuperscript{36} and therefore, a comparison of a given treatment at this point on a patient’s rating of angina, ST segment depression, and oxygen uptake provides an additional efficacy parameter. A close match of the rate-pressure product between treatments is more important than the rate-pressure product per se. We choose a point as late in exercise as possible because treatment effects are often of greater interest closer to the functional limits of the patient.

The gas exchange anaerobic threshold and test end points are discussed in detail below. Needless to say, in any study, heart rate, ST segment responses, blood pressure, rating of perceived effort (or chest pain in angina studies), and gas exchange information are obtained as soon as the test is terminated. The primary question concerning nearly any therapy is its effects on patients’ functional limits.

**Gas Exchange Anaerobic Threshold**

A physiological link between exercise capacity, lactate accumulation in the blood, and respiratory gas exchange was made by Hill and Lupton\textsuperscript{37} more than 60 years ago. A sudden rise in blood lactate level during exercise has long been associated with muscle anaerobiosis and has therefore been termed the “anaerobic threshold.”\textsuperscript{38} Because excess H\textsuperscript{+} ions of lactate must be buffered to maintain physiological pH, CO\textsubscript{2} is produced and ventilation is stimulated. This point of nonlinear increase in ventilation has been used to detect the anaerobic threshold noninvasively and is often termed the “gas exchange anaerobic threshold” (AT\textsubscript{ge}). A great deal of confusion presently exists concerning the mechanism underlying the AT\textsubscript{ge} and how it might be determined and applied clinically.\textsuperscript{39,40}

Changes in the AT\textsubscript{ge} have been used during pharmacological and other investigations to imply a change in oxygen supply to the working muscle, particularly among patients with congestive heart failure.\textsuperscript{20,41–43} It has recently come under scrutiny, however, on the basis of both theoretical\textsuperscript{40} and pragmatic\textsuperscript{44,45} grounds. Connet et al\textsuperscript{46} studied dog gracilis muscle, which is a pure red fiber containing only type I and type IIA fibers, and observed lactate accumulation during fully aerobic, mild (10\% \( \dot{V}\text{O}_{2\text{max}} \)) conditions. These investigators also observed that lactate accumulation was not altered by changes in blood flow, and that lactate accumulation occurred even though no anoxic areas were present in the muscle. This suggests that lactate production and muscle hypoxia are unrelated. Additionally, the advent of tracer technology has raised strong questions about the cause and effect relation between oxygen availability to the muscle and the anaerobic threshold. Many studies now suggest that lactate production occurs at all times, even in resting conditions.\textsuperscript{40} Further, the turnover rate of lactate (the ratio of appearance and disappearance) is linearly related to oxygen uptake during exercise.\textsuperscript{40,47,48} The cumulative effect of these studies has led to the conclusion that the “anaerobic” threshold is not related to muscle anaerobiosis but instead reflects simply an imbalance between lactate appearance and disappearance.\textsuperscript{40}
TABLE 2. Common Testing End Points and Their Limitations

<table>
<thead>
<tr>
<th>Test end point</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age-predicted maximal heart rate</td>
<td>Many regression equations, high variability (SD, 10-15 beats/min)</td>
</tr>
<tr>
<td>Oxygen uptake plateau</td>
<td>Inconsistent criteria, definition, data sampling</td>
</tr>
<tr>
<td>Borg scale&lt;sup&gt;52&lt;/sup&gt;</td>
<td>Subjectivity, reliability depends on subject experience</td>
</tr>
<tr>
<td>Respiratory exchange ratio</td>
<td>High variation in maximal values</td>
</tr>
</tbody>
</table>

It should be noted that although the ATge has been criticized on the basis of studies using tracer technology,<sup>40</sup> the validity of tracer techniques has also been questioned.<sup>49</sup> Thus, the precise mechanism underlying the ATge remains to be delineated. On the basis of recent studies, the following suggestions may be made concerning the use of the ATge for drug studies: 1) Regardless of the mechanism, ventilator changes appear strongly correlated with a lactate threshold, and 2) an alteration in the ATge reflects a change in the balance between lactate production and removal, and references to muscle anaerobiosis should be avoided. Because lactate is strongly associated with muscle fatigue, a change in this relation that can be attributed to pharmacological intervention may add important information concerning the intervention. In this context, the ATge during exercise testing remains an interesting and applicable index for use during pharmacological studies.

An additional consideration concerns the method of choosing the ATge. Our laboratory,<sup>50</sup> in agreement with others,<sup>44,45</sup> has observed that the ATge can vary markedly depending on both the observer and the method of determination. Although a number of methods of determination have been proposed, Caiozzo et al<sup>51</sup> reported that the use of the ventilatory equivalents for oxygen uptake (VE/Vo<sub>2</sub>) and carbon dioxide (VE/VCO<sub>2</sub>) most closely reflected a lactate inflection point. Many laboratories have therefore defined the ATge as the beginning of a systemic increase in VE/Vo<sub>2</sub> without a concomitant increase in VE/VCO<sub>2</sub>. For pharmacological studies, we have successfully used a method outlined by Sullivan et al<sup>4</sup> in which two experienced, blinded (to patient name and whether the test represents a drug or placebo phase) observers independently choose the ATge for each exercise test. When a discrepancy exists, a third observer is also blinded and chooses the ATge independently. The ATge is determined as the minute sample in which two of the three observers agree. The ATge is not included in the analysis for that particular patient when all observers differ. We have found that two observers agree 72% of the time, and two of three observers agree on 100% of tests.<sup>4</sup> In a more recent study, this method resulted in 7% of tests being excluded.<sup>50</sup> This technique avoids interobserver bias and provides a means by which the ATge can be determined objectively.

Testing End Point

Consistent end points for exercise testing are of paramount importance for studying the efficacy of a drug. This is true for both single center studies, generally in which the same patients are compared on drug and placebo, and multicenter studies, in which different patients and doses are compared across centers. In patients not limited by angina, the following testing end points have been used: 1) 80–100% of age-predicted heart rate, 2) a given perceived level of effort by using the Borg scale;<sup>52</sup> 3) a plateau in oxygen uptake, or 4) a given respiratory exchange ratio. All of the commonly used end points have inherent problems; these are outlined in Table 2. A large variance in age-predicted heart rate has been repeatedly observed,<sup>53</sup> with SDs ranging from 10 to 15 beats. Thus, an age-predicted maximal heart rate will be maximal for some patients and submaximal for others, making heart rate an inappropriate end point for studying interventions.

By using gas exchange information, both respiratory exchange ratio (CO<sub>2</sub> production divided by oxygen uptake) (RER) and a plateau in oxygen uptake have been commonly used to suggest that an adequate effort has been observed. An RER of 1.1 or greater indicates that carbon dioxide production exceeds the rate at which oxygen is consumed. The source of this excess carbon dioxide has long been thought to represent the bicarbonate necessary to buffer lactate, the production of which increases greatly as exercise progresses. Although an RER greater than 1.1 suggests a good patient effort, maximal RER varies greatly from patient to patient and therefore is a poor cut point to identify maximal effort.

A plateau in oxygen uptake (the failure of oxygen uptake to increase with increasing work) has long been considered the classic marker for an individual’s cardiopulmonary limit. The plateau concept, however, has suffered from inconsistencies in definition, criteria, and data sampling, all of which affect the interpretation of the relation between changes in work and oxygen uptake. Our laboratory recently noted that in the same patients, a plateau may or may not occur depending on the application of different criteria.<sup>54</sup> Additionally, the sampling interval used has a pronounced effect on the occurrence of a plateau, and the observation of a plateau may vary from day to day. This may explain the fact that a
plateau has been reported to occur in between 7% and 75% of individuals tested.\textsuperscript{54,55} Despite its wide application for over 30 years, the plateau remains an inconsistent and somewhat tenuous concept. Thus, its use for drug studies in cardiovascular disease is inappropriate.

Because of the limitations of heart rate and oxygen uptake to detect maximal effort, it is important to quantify the patient’s perception of effort. Although a quantification of the patient’s perceived effort during exercise is subjective, we have found the Borg scale\textsuperscript{52} to be the most useful method to assure maximal effort. Close correlations have been observed with a number of physiological parameters during exercise.\textsuperscript{56} Together with other laboratories, we have observed that ratings of perceived effort may be most closely related to the sensation of breathing.\textsuperscript{57} We rate the patient’s effort every minute and at peak exercise, although most laboratories rate perceived exertion at the end of each stage. Like familiarization with the treadmill, bicycle, or gas exchange equipment, ratings of perceived exertion are more reliable when patients have had sufficient experience, and this is an additional application of the prestudy screening exercise test.

In angina studies, the effect of the intervention on the patient’s perception of chest pain during the test is usually the primary efficacy parameter. The description of the exercise end point in these studies varies from laboratory to laboratory, ranging from “maximal tolerated angina” to “onset of angina.” Of utmost clinical importance, however, is the influence the drug has on the patient’s activities of daily living or need for sublingual nitroglycerin. Therefore, the end point of these exercise tests should be the point of pain that would typically cause the patient to stop his activity or take nitroglycerin outside the laboratory setting. Stopping the exercise test short of this point underestimates the drug effect, and asking the patient to exercise beyond this point is a disservice to the patient. We have modified a 1-to-4 scale outlined previously\textsuperscript{58} to describe the patient’s symptoms during exercise (1, onset of chest discomfort; 2, moderate discomfort; 3, moderately severe discomfort; and 4, severe discomfort). A rating of 3, or “moderately severe” is the test end point, and represents the point at which the patient would stop and rest, or take a sublingual nitroglycerin pill to relieve the pain, or both. We are careful not to coach the patient and make certain that the patient knows beforehand that he or she dictates the duration of the test (subject to the physician’s discretion). When gas exchange apparatus is used, these become simple hand signals. Because angina studies are an evaluation of the degree of chest pain and not overall perceived effort, we do not use the Borg scale\textsuperscript{52} for these studies. This avoids any confusion associated with the use of two scales.

**Conclusion**

An abundance of information regarding medications commonly prescribed in clinical cardiology has evolved from exercise test studies. Even with great attention to detail, the study of treatment effects during exercise testing in heart disease is imprecise. An inappropriate exercise protocol, exercise end point, study design, or the misuse of gas exchange data can increase the chance of making type 1 or type 2 errors. Greater discussion of these issues can only lead to an improvement in the quality of these studies. It is hoped that consideration of these issues will lead to a reduction in extraneous factors that influence study results. The following recommendations are made for optimizing exercise testing when performing drug studies:

1. Patients whose disease is stable and who respond reproducibly (±10%) to serial maximal exercise testing should be selected for studies.
2. Serial testing should be performed on separate days, at the same time of day. 3. The exercise protocol should be individualized for the patient and the purpose of the study. Generally, protocols with small increments in work, with work rates set to yield a test duration of approximately 10 minutes, are recommended. 4. The use of gas exchange techniques greatly improves the precision of cardiopulmonary assessment. 5. Comparing hemodynamic and gas exchange data submaximally (matched rate-pressure product, matched workload, onset of angina, and gas exchange anaerobic threshold), in addition to resting and peak exercise data, can yield important information concerning drug efficacy. 6. The gas exchange anaerobic threshold should be considered a general and not a precise marker of a physiological event. This point does not necessarily or precisely reflect muscle anaerobiosis. 7. Gas exchange data (respiratory exchange ratio and oxygen uptake plateau) and age-predicted maximal heart rate are limited in value as exercise testing end points. There is no physiological standard for a patient’s cardiopulmonary limits. The Borg\textsuperscript{52} and angina\textsuperscript{58} scales are recommended to quantify testing end points.

**References**

12. Wicks JR, Sutton JR, Oldridge NB, Jones NL: Comparison of the electrocardiographic changes induced by maximum exercise testing with treadmill and cycle ergometer. *Circulation* 1978;57:1066–1069
29. Goldman L, Hashimoto B, Cook EF, Losalzo A: Comparative reproducibility and validity of systems for assessing cardiovas-

**KEY WORDS** • exercise testing • oxygen uptake • anaerobic threshold • heart failure • angina
Optimizing the exercise test for pharmacological investigations.
J Myers and V F Froelicher

Circulation. 1990;82:1839-1846
doi: 10.1161/01.CIR.82.5.1839

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/82/5/1839