In Clinical Trials, Is the 6-Minute Walk Test a Better Functional Test of Interventions for Peripheral Artery Disease Than Treadmill Walking Tests?

The Treadmill Is a Better Functional Test Than the 6-Minute Walk Test in Therapeutic Trials of Patients With Peripheral Artery Disease

William R. Hiatt, MD; R. Kevin Rogers, MD; Eric P. Brass, MD, PhD

Patients with peripheral artery disease (PAD) have a severe functional limitation in community walking ability, peak exercise performance, and activities of daily living. The limb manifestations of PAD that are associated with reduced exercise performance not only are classic intermittent claudication but also include atypical symptoms ranging from leg discomfort during exercise to exercise limitations without any clearly referable symptoms in the lower extremity. Despite the diverse nature of limb symptoms during exercise, patients with PAD uniformly have a severe reduction in peak and endurance exercise performance. Thus, in this article, the term PAD-limited is used to define the exercise impairment that is well described in PAD, replacing the term claudication-limited to better characterize the diverse nature of exercise-limiting symptoms described in the disease.

Response by McDermott et al p 78

The quantification of the exercise limitation in PAD and the implications for daily walking ability and community activities can be conceptualized by understanding the metabolic work required to perform a range of activities. The metabolic equivalent of task (MET) is based on a multiple of the resting metabolic rate (≈3.5 mL oxygen-min⁻¹·kg⁻¹) and can be used to characterize a given activity. The MET can be used to rank physical activities from very mild (walking at a slow pace is 2 METs) to intense (running at 6 mph is 10 METs). Common activities such as walking on level ground at a speed >3 mph, vacuuming, shopping for groceries, and transporting packages all require 4 to 5 METs. Measurements of oxygen consumption can also be made to characterize both submaximal and maximal workloads and related to METs. Typically, patients with PAD have a PAD-limited peak exercise performance of ≈15 mL oxygen-min⁻¹·kg⁻¹ or the equivalent of 4 to 5 METs, which is half that expected in an age-matched healthy control subject. Thus, limitations in peak exercise performance can translate directly to an inability to complete common activities in the PAD population. Furthermore, the higher the percentage of maximal capacity required to perform a daily task is, the harder the task will seem to the patient and the shorter the patient will be able to engage in the activity. Practically, this is equivalent to patient-perceived endurance while performing work and is dependent on the intensity of the specific activity. As a result, characterization of peak

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.
From the University of Colorado School of Medicine, Department of Medicine, Division of Cardiology and CPC Clinical Research, Aurora (W.R.H., R.K.R.); and Harbor-UCLA Center for Clinical Pharmacology, Torrance, CA (E.P.B.).
This article is Part II of a 2-part article. Part I appears on p 61.

Correspondence to William R. Hiatt, MD, Professor of Medicine/Cardiology, University of Colorado School of Medicine, C/O CPC Clinical Research, 13199 E Montview Blvd, Ste 200, Aurora, CO 80045. E-mail Will.Hiatt@UCDenver.edu
(Circulation. 2014;130:69-78.)
© 2014 American Heart Association, Inc.

Circulation is available at http://circ.ahajournals.org
DOI: 10.1161/CIRCULATIONAHA.113.007003
exercise performance can be used to assess a patient’s ability to conduct a variety of normal ambulatory activities, both in an absolute sense (can or cannot do) and as related to endurance or perceived exertion for a given task.

The primary treatment goal in ameliorating symptoms of PAD, independently of the type of intervention (eg, revascularization, exercise training or pharmacological therapy), is to improve the patient’s ambulatory function and quality of life. The effectiveness of an intervention can be understood on the basis of improvements in peak and submaximal exercise capacity because these in turn are determinants of ambulatory function. This improvement can best be assessed by direct, formal exercise testing. Surrogates of clinical benefit used to assess the results of a revascularization, including changes in hemodynamics such as the ankle-brachial index or changes in vascular anatomy as assessed by vascular imaging such as duplex ultrasound, magnetic resonance, or computed tomography imaging, do not directly provide insight into the resultant functional improvements. In contrast, an exercise test is a direct surrogate for what matters to the patient: improved physical functioning and ability to perform and sustain activities important to that patient’s quality of life.

**Pathophysiology of the Exercise Limitation in PAD**

In healthy subjects, maximal exercise capacity reflects the integrated physiological limitation of central hemodynamics (ability of the lungs to allow the transfer of oxygen to the blood and the ability of the heart to deliver blood to the lower extremities), the ability of the muscle to extract oxygen and use it for energy production, and the ability of the muscle to translate chemical energy into mechanical work. Each component of this system is essential, and changes in any one can result in a change in peak exercise performance. For example, exercise training in healthy subjects induces an array of adaptations, including an increase in muscle mitochondrial content and thus an improvement in the ability to generate energy with the delivered oxygen. In disease states, impairment in any component will result in a physiological limitation reflected in reduced exercise capacity such as in heart failure or lung disease that affect central hemodynamics.

Similarly, in PAD, the complex insult of chronic ischemia results in changes in both blood delivery and efficient use of the oxygen by skeletal muscle, leading to a unique pathophysiological limitation. Patients with PAD have a limitation in blood flow to the skeletal muscle, resulting in a mismatch of oxygen supply to meet metabolic demand. In the absence of critical limb ischemia, symptoms occur only during exercise because blood delivery is adequate to meet resting metabolic requirements. However, inadequate blood flow and oxygen delivery during exercise do not fully define the functional deficit in PAD. There are secondary sequelae to the perfusion defect that lead to diverse changes to the distal perfusion bed, including to the microvasculature and muscle function. Chronic oxidant stress may be a common injury mechanism in PAD. These secondary injury phenomena include mitochondrial dysfunction such that oxygen consumption is impaired even when oxygen delivery is adequate at the onset of exercise.

Further evidence of the complex pathophysiology of claudication is the observation that exercise training results in a substantial increase in peak walking time (PWT), which is often greater than the increase in peak oxygen consumption, despite correlation in the change between the 2 parameters. The improvement in peak oxygen consumption is consistent with a classic training response, but other factors such as improved bioenergetics (increase in ATP generated per oxygen consumed) may also contribute to the improved performance in patients with PAD after training. These exercise improvements are not correlated with changes in limb hemodynamics but may be associated with changes in skeletal muscle metabolism and capillary density. In contrast, revascularization results in substantial improvements in limb blood flow and oxygen delivery, but these improvements in limb flow are not well correlated with the change in exercise performance.

The recognition that a limitation in maximal exercise performance is a direct result of a complex pathophysiology and that both endurance and submaximal exercise performance are inherently impaired when maximal performance is limited has important implications for considering how to assess function and response to treatment in these patients.

**Optimal Functional Assessment of Interventions for PAD**

Several characteristics of an optimal functional test can be postulated to assess the results of treatment of or an intervention in PAD. This is particularly important when the test serves as a primary or secondary end point in a clinical trial of a new treatment used for regulatory approval. Table 1 summarizes several key concepts in functional testing. As noted, patients with PAD have a severe limitation in functional capacity related to a complex pathophysiology. The optimal test should directly reflect the patient’s limitation and the underlying disease characteristics. Ideally, the functional test should be able to assess the impact of an intervention even when the mechanism of benefit is not fully understood (eg, cilostazol).

Patients with PAD have a wide range of functional limitations, from severe challenges in ambulation (unable to walk 50 m) to minimal physical restrictions (able to walk 1 km but limited when jogging). Therefore an optimal test should have acceptable performance characteristics across a wide dynamic range of exercise limitations typical of the disease population. In the context of a clinical trial, the baseline performance should be reliably established to assess changes in performance with an intervention. Furthermore, when patients undergo repeat testing over a period of time in the absence of an intervention, the test...
should provide reproducible results. As a result, in the context of a clinical trial, patients randomized to a placebo group should optimally have stable responses over time.

Critically, any change in test performance should be directly correlated with a patient’s perception of his or her limitation and change in ambulatory functional status. Data supporting this relationship between laboratory test performance and patients’ outpatient, real-world function are critical to evaluating the utility of the laboratory measurement. In this context, the exercise test is really a surrogate for the primary outcome of interest, that is, improvement in the patient’s functional status associated with a wider range of physical activities. In addition, a quantified unit of change on the test should be interpretable in the context of a minimum detectable difference, which is related to the test variance, and minimum clinically important difference, which is the smallest change in the exercise test that the patient would report as clinically important.6,25,26

When used to assess the results of an intervention, all functional tests have inherent variability. Typically, there are 2 key factors: the actual patient variability in the response to a treatment and the assay reproducibility of the test. The reproducibility of the test can be described by the repeat test coefficient of variance in which the same subject undergoes repeat testing under stable clinical conditions over a period of days or weeks. However, in the context of an intervention, the standardized effect size is a useful way to relate the ability to detect change on the basis of the heterogeneity of the studied population. The standardized effect size is the difference in treatment responses between the active and control arms of a trial normalized to the pooled standard deviation of the test (variance of the control and experimental groups combined).27 The population variability (pooled standard deviation) reflects both the test variability and the response variability. Thus, if a treatment has a very large effect (eg, exercise training), the sample size to demonstrate a statistically significant difference between groups can be relatively small if the variance in the response across the population is small. A larger variance will result in a larger sample size to achieve statistical significance. In contrast, if the treatment effect is relatively small (eg, drug therapy), then controlling the variability and pooled standard deviation of the test is critically important to identify a statistically significant result with a limited sample size. The standardized effect size can be divided into treatments with low, medium, and large effects, but this approach to grading the magnitude of response on a treadmill (or other functional test) may not relate the degree of improvement in the test to what the patient would regard as important in his or her daily activity.

Any functional test performed in PAD should be safe and well tolerated. The test should have minimal risk in this patient population with underlying cardiovascular disease and increased risk of cardiovascular events.28 The burden of the test should be acceptable for patients because in a clinical trial the test will be repeated over time. Equipment for the test should be readily available and should be standardized across study sites when used in the context of a multicenter clinical trial. The cost of equipment and performance of an individual test should be reasonable. Finally, personnel at the study sites will typically need to be trained in conducting a standardized and reproducible test. The testing procedures should be acceptable to study sites in a clinical trial.

### Table 1. Characteristics of an Optimal Functional Test in Clinical Trials for PAD

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Optimal Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characterizes functional limitation associated with disease state</td>
<td>Test reflects underlying disease pathophysiology</td>
</tr>
<tr>
<td>Large dynamic range</td>
<td>Reliably assesses broad range of functional limitations in the disease population</td>
</tr>
<tr>
<td>Baseline</td>
<td>Baseline value clearly established</td>
</tr>
<tr>
<td>Reproducible</td>
<td>Provides stable results over time in the absence of an intervention or disease progression</td>
</tr>
<tr>
<td>Correlates with patient-reported outcomes</td>
<td>Directly correlates with patient-reported functional status and physical limitations</td>
</tr>
<tr>
<td>Sensitive to detecting treatment effect</td>
<td>Detects a change in function with an active treatment with acceptable effect size</td>
</tr>
<tr>
<td>Established minimum clinically important difference</td>
<td>Pooled standard deviation of the test is small relative to the treatment effect</td>
</tr>
<tr>
<td>Safe and well tolerated</td>
<td>Has implications for sample size calculations</td>
</tr>
<tr>
<td>Used in multicenter trials</td>
<td>Used to differentiate a statistical difference from a clinically meaningful difference.</td>
</tr>
<tr>
<td>Cost</td>
<td>Has low risk of cardiovascular or orthopedic events</td>
</tr>
<tr>
<td>Equipment available at a wide range of sites</td>
<td>Is acceptable and not burdensome to patients</td>
</tr>
<tr>
<td>Site training produces consistent results across centers</td>
<td>Extensive experience in performing the test across diverse countries and study sites</td>
</tr>
<tr>
<td>Equipment available at a wide range of sites</td>
<td>Acceptable costs of equipment and personnel to conduct the test</td>
</tr>
</tbody>
</table>

PAD indicates peripheral artery disease.

### Treadmill Exercise Testing

#### Methodologies in PAD

**PAD Treadmill Protocols**

Standardized methods for treadmill exercise testing have been developed for a variety of populations.29 In PAD, there...
are 2 basic treadmill exercise protocols: the constant-load test and the graded test. The constant-load test is performed at a single work rate, typically 2 mph (3.2 km/h) at a grade of 10% to 12%. The constant-load protocol evolved out of the vascular laboratory many years ago where an exercise test was used to provoke a drop in ankle pressure to facilitate the diagnosis in patients with mild PAD. The results of a constant-load test are often expressed in units of distance walked, but this approach does not permit an accurate assessment across the range of functional impairment in the broad population of PAD subjects. For example, the often used workload of 10% to 12% grade is quite extreme and often exceeds a patient’s individual functional limitation, resulting in an inability to conduct the test. In contrast, some patients with mild PAD will be able to walk for an extended period on the constant-load protocol without being limited by claudication and thus precluding assessment of their maximal limitation. Thus, only a limited spectrum of the PAD population can be assessed with the constant-workload treadmill test.

In contrast, the most common graded test used in PAD begins the workload at 2 mph, 0% grade. This work rate is typical of low-intensity (=2 METs) activities of daily living and is tolerated by the vast majority of patients with symptomatic PAD. Every 2 minutes, the grade is increased 2% until the patient is limited by maximally tolerated claudication pain and can walk no further. If a patient with PAD does not have classic claudication, the same test can be done. In this situation, the patient is limited by his or her atypical or nonspecific symptoms, but nonetheless premature, muscle fatigue. With a progressive increase in the workload, each patient is taken to an individually defined, PAD-limited peak exercise performance so that even patients with mild PAD will reach a PAD-limited workload. The graded treadmill test has a large dynamic range that can reproducibly define an individual patient’s peak PAD-limited PWT and claudication onset time that occurs before peak workloads. In addition, measurements of oxygen consumption have been used to validate the PWT and to reflect the underlying pathophysiology of altered oxygen delivery and consumption. In research programs, measurement of ventilatory gas exchange can be included in the test, but this requires more specialized equipment, and is more demanding of subjects, and thus is not practical in most clinical trials.

The use of graded exercise testing is well established and accepted in clinical medicine. Graded testing has been used in clinical cardiology for decades to help evaluate cardiac ischemia and functional limitations in patients with coronary artery disease. The philosophy of increasing workload to create a metabolic imbalance in heart muscle is at the core of the Bruce protocol and other graded tests in cardiology and is directly analogous to the situation in the lower limb of patients with PAD. This history of extensive previous use also means that high-quality graded treadmills are widely available. Thus, PWT on a graded treadmill test represents a true physiological maximal performance based on the individual patient’s limb pathophysiology that can easily be performed in diverse clinical settings.

**Conduct of the Exercise Test**

The utility of any exercise test is dependent on proper implementation. Thus, it is important that the graded treadmill exercise test be performed under standardized and reproducible conditions. Typical instructions are to have patients tested in the morning after an overnight fast so that food effects can be minimized. The patient should not walk long distances to the testing laboratory (to avoid claudication), be rested, and should avoid smoking cigarettes and drinking alcohol before the test. Comfortable clothes should be worn, and the timing device used to assess test duration should be hidden from the patient so as not to bias their performance. The patients should initiate the test by straddling the moving belt at 2 mph and start walking at that speed at time 0. This will require familiarization of all patients before the conduct of a formal test for analysis. During the conduct of the test, patients should not hold onto the treadmill bars for support or off-loading because this will alter the subject’s performance and introduce variability in the test. Testing laboratories should consistently instruct patients to walk to maximally tolerated claudication or other limb pain/limitation. For example, some investigators instruct their patients to stop walking on the treadmill when they experience a usual intensity of claudication or other limb pain. However, these instructions will, by definition, artificially limit that patient’s peak performance below their true pathophysiological maximal limitation. This approach will result in data that are less reproducible and less sensitive to the intervention with less predictive value with respect to ambulatory performance. Thus, all patients should be pushed to maximally tolerated pain that results in an inability to walk any further. Although the safety of treadmill testing has not been evaluated extensively in the PAD population, 1 study included comprehensive hemodynamic and ECG monitoring during exercise in a large cohort and identified no safety concerns, including no changes indicative of cardiac ischemia during testing. This finding is consistent with the reported adverse event experience from clinical trials using peak treadmill testing involving thousands of patients with PAD. However, most patients with PAD have underlying coronary artery disease; therefore, ECG monitoring during all peak treadmill exercise tests in this population is recommended to enhance patient safety.

**Challenges in Performing a Graded Treadmill Test in a Clinical Trial**

There are several challenges in performing a graded treadmill test across multiple sites in a clinical trial. Not all sites maintain appropriate equipment; some treadmills are simply
unable to perform the test according to the study protocol.\textsuperscript{38} For example, treadmills may not be programmed correctly, may not increment the grade correctly, or may rely on manual changes in grade that are often subject to error. Treadmill speed must be calibrated and stable over the entire testing range. These issues need to be identified and corrected before a site enrolls patients in a clinical trial. In addition, site personnel and study subjects may not be fully trained on the proper conduct of the test. All subjects, particularly patients with a chronic disease, must be familiarized with the testing procedures before performing an exercise test for data analysis. In addition, site staff should use standardized methods for initiating the test (proper transition from rest to 2 mph, 0\% grade) and completing the test by pushing the patient to a maximal, PAD-limited performance.

Thus, assessing the treadmill testing equipment and staff competency at a site is critically important to achieve uniform treadmill end points of claudication onset time and PWT across a multicenter trial. Routine clinical trial site regulatory monitoring usually does not involve an evaluation of the actual conduct of the test. Several methods have been proposed to ensure that study sites are properly trained and comply with standardized testing procedures. One approach to this standardization is the site end-point evaluation visit, which is an extension of the usual monitoring visit. The site end-point evaluation visit evaluates the capability and equipment of the site to perform a test and observes the performance of the test at the site using mock testing methodologies. This allows further training of the site staff to provide a uniform approach across sites to reproducibly perform the exercise test. Using these methodologies and defining other aspects of test performance have provided much greater control of the variance of the exercise test between and within individuals and sites in the conduct of a clinical trial. Of note, this approach to control of testing conditions is likely applicable to all exercise testing modalities, including the 6-minute walk test.

**Treadmill Reliability**

The test characteristics of treadmill testing have been studied extensively in PAD. For example, numerous studies have evaluated the reliability of the constant-load compared with the graded treadmill test. A meta-regression analysis of 8 studies compared the 2 tests using the intraclass correlation coefficient to determine the consistency between measurements over time.\textsuperscript{39} It should also be noted that these 2 exercise tests have different relationships between workload and PWT. The constant-load test has a linear relationship, whereas the graded test has a curvilinear relationship (higher work rates are associated with progressively shorter incremental walking times).\textsuperscript{39} The reliability of PWT was significantly better with the graded test (intraclass correlation coefficient, 0.95) compared with the constant-load test (intraclass correlation coefficient, 0.90). Consistent with the limitations enumerated above, the reliability of the constant-load test was also dependent on the fixed grade of the treadmill (poor at 0\% grade and better at 12\% grade), whereas the graded test performed well across the full dynamic range of patients studied. Other studies have also confirmed better reproducibility with the graded compared with the constant-load test in PAD.\textsuperscript{40} Importantly, in evaluations of patients with PAD, the high baseline reproducibility of the graded treadmill has been confirmed with the use of data from a multicenter trial.\textsuperscript{41} The intrasubject coefficient of variation averaged 15\% and was independent of the baseline exercise performance (the test was equally reproducible across the full range of baseline PWT values). Thus, although both treadmill protocols are acceptable, the graded test better reflects the mechanism of the performance limitations in PAD, has the greatest reliability, and allows testing a wide range of functional disease severity in PAD.

A major challenge of any treadmill test in PAD is an apparent improvement in peak or submaximal performance measures over time in the absence of an intervention. This phenomenon has also been called the placebo response because of the increase in performance observed in patients randomized to a placebo group over 6 months of observation.\textsuperscript{31} In prior studies using a constant-load exercise test, the placebo response has varied from 25\% to 100\%.\textsuperscript{42,43} However, contemporary multicenter trials using the graded treadmill test and site training (site end-point evaluation visit) have controlled this placebo response to 9\% to 12\% at 3 months and 13\% to 23\% at 6 months.\textsuperscript{44} The cause of this placebo response is likely multifactorial. The subject’s comfort and adjustment to walking on the treadmill, including optimization of gait, likely contribute. In principle, if the maximum performance is properly established, variability over time should be very small because of the fixed physiological limitation. The observed changes are also consistent with a change in walking efficiency as a contributing factor.

An additional factor affecting the variance of any exercise test in a clinical trial is the definition of the baseline performance before randomization to an intervention. A recent study carefully evaluated the definition of baseline exercise performance for the graded treadmill test. In the context of a multicenter clinical trial, patients performed 3 consecutive treadmill tests 3 to 10 days apart before randomization to treatment with an investigational drug, an active comparator drug, or placebo.\textsuperscript{41} The data were evaluated by assessing various combinations of baseline definitions. The methods that resulted in the lowest placebo response and the largest standardized effect size were defining baseline either by the first of the 3 tests or by the highest of the 3 tests, whereas the characteristics of averaging the 3 tests or any set of 2 tests were inferior.\textsuperscript{41} Using the highest baseline value (which we recommend) supports the concepts that the patient cannot exceed his or her true pathophysiologic limitation on a graded treadmill test and that this limit will be achieved with proper test execution.

**Treadmill Sensitivity to Treatment Effect**

When used as a primary end point, PWT assessed with a graded treadmill test is sensitive to change with a variety
of interventions. Statistically significant responses relative to a control group have been documented for exercise rehabilitation, revascularization, and a variety of medications. The treadmill has been used in several drug trials of agents, including those with small effect sizes. For example, a meta-analysis of propionyl-L-carnitine demonstrated an effect size in a range of 0.11 to 0.22, whereas phosphodies- terase-3 inhibitors have effect size estimates in the range of 0.28 to 0.32. Indeed, no intervention demonstrated to be effective has failed to demonstrate improvement in 1 or more clinical trials using graded treadmill testing.

The Treadmill as a Surrogate for Patient Ambulatory Function and Quality of Life

Treadmill performance has been related to several measures of ambulatory activity. Performance on the graded treadmill is highly correlated with the Walking Impairment Questionnaire, which is a patient-based assessment of challenges of walking defined distances, walking speeds, and severity of calf claudication pain during community walking activities. The Walking Impairment Questionnaire score can range from 0.00 to 1.00 (or 0% to 100%), with higher scores indicating better walking ability. There are also subscales for walking distance, speed, stairs, and claudication pain severity. The relationship between improvements in treadmill performance and change in Walking Impairment Questionnaire scores was further evaluated in a study of exercise training. This study demonstrated that a change of 0.1 in Walking Impairment Questionnaire score (mean of the subscales) corresponded to a 345-m change in peak treadmill walking distance with exercise training. Although the minimum clinically important difference was not formally assessed, there was a suggestion that different quartiles in treadmill responses could identify potential thresholds for patient-assessed improvements in walking impairment and functional status. In addition, patient-reported outcomes have been assessed in trials of exercise training, drugs, and revascularization. In all these studies of a variety of interventions, changes in PWT on a graded treadmill test were well correlated with improvements in patient-assessed outcomes and health-related quality of life.

Maximal walking distance on a progressive treadmill protocol was also correlated with outdoor walking distance measured with a global positioning system. This relationship is relatively weak ($r^2=0.28$), indicating that peak performance on a graded treadmill may reflect different pathophysiological limitations than does a submaximal effort with the 6-minute walk test.

Assessment of Performance and Response to Treatment With the Treadmill Versus the 6-Minute Walk Test

The 6-minute walk test is an attractive assessment modality on many levels. It is simple to perform, requires minimal equipment, and has been used in a variety of disease states, including in patients with pulmonary disease and heart failure, but less so as an end point in PAD. In PAD, although the graded treadmill test and 6-minute walk test are correlated, the relationship is relatively weak ($r^2=0.28$), indicating that peak performance on a graded treadmill may reflect different pathophysiological limitations than does a submaximal effort with the 6-minute walk test.

In contrast to the extensive experience with the treadmill, equivalent data for performance of the 6-minute walk test in patients with PAD are not available. However, data from other disease populations suggest that there are several factors in the conduct of the 6-minute walk test that will affect outcome. For example, in patients with chronic lung disease, course length can influence the measurement. As with a treadmill test, precision of the 6-minute walk test is improved with the use of the higher of 2 baseline measurements and with test familiarization.

In chronic lung disease, a minimal clinically important difference has been defined for the 6-minute walk test predicting mortality (a loss of 30 m). In contrast, in patients with heart disease, the 6-minute walk test demonstrated poor test-retest reliability, as well as a placebo response and time-of-day variability. A minimum clinically important difference has been suggested in chronic lung disease and in heart failure that defined thresholds for change in a patient-reported outcome scale, but in heart failure, this difference was the same as the variance of the test and hence the minimum detectable difference. Thus, the test would be of marginal utility in assessing an intervention with an effect size that approximated the minimum clinically important difference.

When used in PAD, the 6-minute walk test has demonstrated a loss of patient performance over several years of follow-up that is related to a variety of clinical and systemic factors such as level of chronic inflammation. However, any relationship between this loss of walking performance over time and patient assessment of changes in quality of life was not assessed. Similar data on functional decline have not been obtained with the treadmill test, but this particular outcome (preventing clinical disease progression) has not been the target of interventions tested in PAD randomized, clinical trials. In addition, when evaluated in the context of an intervention, an endurance test such as the 6-minute walk has no fixed physiological correlate to support a change in function because the test is performed under submaximal exercise conditions. Thus, performance (6-minute distance walked) may be susceptible to environmental and intervention-independent factors. Additionally, patient and investigator expectations of benefit may affect the submaximal test stability over time. This is particularly important in the context of a randomized trial in which blinding to the intervention may be impossible.
(eg, exercise training) or compromised by factors such as drug adverse events. In contrast to the abundant data validating treadmill performance as a surrogate of ambulatory patient function, there are fewer data exploring this relationship for the 6-minute walk test in PAD subjects.1

In PAD, very few studies have offered a head-to-head comparison of the 6-minute walk with graded treadmill testing to evaluate an intervention, but 2 studies have been published by McDermott et al.67,68 In 1 single-site study, the response to supervised exercise training over 6 months was evaluated with both tests.67 In the control group, both measures were relatively stable over 6 months, with the 6-minute walk test demonstrating a loss of walking distance of 5% the graded treadmill demonstrating a 7.6% increase. With the exercise intervention, the 6-minute walking distance improved 6.4% (P<0.001 versus baseline), whereas the graded treadmill test improved 52% (P<0.001 versus baseline). Variance estimates of the within-group changes in the exercise group for both tests were determined from data provided in the article, leading to a calculation of a standardized effect size (difference in treatment response between groups divided by the within-subject standard deviation). The between-group comparisons of the exercise and control group changes were statistically significant for both the 6-minute walk test and the treadmill test. On the basis of those changes and data from the publication, the 6-minute walk test had an estimated effect size of 0.70 compared with an estimate of 1.01 for the graded treadmill. These differences in effect size evaluating the same intervention have implications for future trials in which the graded treadmill test (greater observed effect size) would require fewer patients to demonstrate statistical significance than the 6-minute walk test. The other single-site study evaluated home-based exercise and suggested that the estimate of the treatment effect (effect size) was 1.4 times greater with the graded treadmill test compared with the 6-minute walk test.68 Taking these results taken together and combining them with the characterization of each test in PAD and other conditions indicate that the use of the 6-minute walk test as a primary end point in assessment of an intervention would require a larger sample size than if the graded treadmill were used. In contrast to the substantial multicenter experience with treadmill testing in a real-world setting, the above data are from a single-site study. Experience suggests that single-site test performance should be extrapolated with caution to multicenter trials. These issues, including experience with the treadmill and 6-minute walk test in multicenter trials, are summarized in Table 2.69

### Table 2. Characterization of the Exercise Treadmill Test Compared With the 6-Minute Walk Test in Clinical Trials in PAD

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Graded Treadmill</th>
<th>6-Minute Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characterizes disease functional limitation</td>
<td>Results informative for both submaximal and maximal performance</td>
<td>Characterizes submaximal walking endurance</td>
</tr>
<tr>
<td>Dynamic range</td>
<td>Accurate assessment over a wide range of exercise limitations</td>
<td>Less well studied in PAD</td>
</tr>
<tr>
<td>Baseline</td>
<td>Highest value of a series of tests at baseline facilitates best estimate of treatment effect41</td>
<td>Not established in PAD</td>
</tr>
<tr>
<td>Experience</td>
<td>Assessed in multicenter trials</td>
<td>In pulmonary disease, higher of 2 tests recommended68</td>
</tr>
<tr>
<td>Reproducible</td>
<td>“Placebo” response controlled with proper conduct of the test67</td>
<td>Controlled with proper conduct of the test67</td>
</tr>
<tr>
<td>Variability</td>
<td>Coefficient of variation 12.6%47</td>
<td>Coefficient of variation 10.4%47</td>
</tr>
<tr>
<td>Correlation of performance parameter with patient-reported outcomes after treatment</td>
<td>Correlations observed in studies of drugs, exercise, and revascularization23,25,26,46</td>
<td>Correlation not evaluated but improvement seen in exercise studies41,48</td>
</tr>
<tr>
<td>Sensitive to treatment effect</td>
<td>Effect size, medication 0.11–0.3244,47</td>
<td>Not well studied</td>
</tr>
<tr>
<td></td>
<td>Effect size, exercise training 0.87–1.0144,45,47</td>
<td>0.7047</td>
</tr>
<tr>
<td></td>
<td>Effect size, revascularization 0.4746</td>
<td>Not well studied</td>
</tr>
<tr>
<td>Minimum clinically important difference</td>
<td>Not established in PAD</td>
<td>Not established in PAD; thresholds suggested in chronic lung disease and heart failure64,65,66</td>
</tr>
<tr>
<td>Safe and well tolerated</td>
<td>Extensive experience without unacceptable dropouts resulting from adverse events Cardiac safety well established in PAD population47</td>
<td>Not well studied but test is less burdensome than a treadmill</td>
</tr>
<tr>
<td>Multicenter clinical trials</td>
<td>Broad experience in diverse sites across the world83 Study sites require training</td>
<td>Most experience limited to single-site studies Likely more acceptable to diverse sites than treadmill testing</td>
</tr>
<tr>
<td>Costs</td>
<td>Requires a programmable treadmill and ECG monitoring</td>
<td>Requires no specialized equipment</td>
</tr>
</tbody>
</table>

PAD indicates peripheral artery disease.
Conclusions
The primary goal of any intervention for symptomatic PAD is to improve the patient’s ambulatory performance of daily activities and quality of life. The 6-minute walk test may have minor advantages over treadmill testing with respect to study site and patient preferences, considerations that deserve further evaluation. Its use for other indications also suggests that it is a useful measure of physical performance. Nonetheless, the theoretical advantages of the graded treadmill test have been widely confirmed in >30 years of clinical research. It has a sound physiological basis, has been widely used in multicenter trials, and has broad acceptance in clinical practice for other conditions. In PAD, the graded treadmill intrasubject and intersubject test characteristics are well established and support the utility of the test while facilitating sound power estimates for clinical trials. In contrast to the established and support the utility of the test while facilitating sound power estimates for clinical trials. In contrast to the established and support the utility of the test while facilitating sound power estimates for clinical trials. In contrast to the established and support the utility of the test while facilitating sound power estimates for clinical trials.

Disclosures
Dr Hiatt reports grant awards to CPC Clinical Research (a nonprofit academic research organization and affiliate of the University of Colorado) from the following sponsors: Aastrom, AstraZeneca, National Institutes of Health, CSL, Cytokinetics, DNAVEC, Kowa, Kyushu University, Pluristem, Regeneron, Riget, Takeda. Dr Rogers reports no conflicts. Dr Brass reports being a consultant to GlaxoSmithKline, Novartis, McNeil Consumer Pharmaceuticals, Novo Nordisk, 3D Communications, Catabasis Pharmaceuticals, Allergan, NovaDigm Therapeutics, NanoBio, Bayer, Endo Pharmaceuticals, Amgen, Boston Scientific, World Self-Medication Institute, Merck, NPS Pharmaceuticals, HeartWare International, Takeda Pharmaceuticals, Genzyme, Consumer Healthcare Products Association, DepoMed, Cangene, Aveo Oncology, BioMarin, Galderma, QRx Pharma, University of Washington, Amarin, EnteroMedics, and Trius Therapeutics. Dr Brass has equity in Calistoga Pharmaceuticals and Catabasis Pharmaceuticals.

References


Response to Hiatt et al

Mary M. McDermott, MD; Jack M. Guralnik, MD, PhD; Michael H. Criqui, MD, MPH; Kiang Lui, PhD; Melina R. Kibbe, MD; Luigi Ferrucci, MD, PhD

We agree with Hiatt et al that the primary goal of therapeutic interventions for patients with peripheral artery disease (PAD) is to improve physical functioning and to increase engagement in activities important to the PAD patient’s quality of life. Among PAD patients, growing evidence demonstrates that the 6-minute walk more directly measures outcomes that are important and relevant to daily functioning and quality of life than treadmill testing. Treadmill walking is an artificial form of walking, requiring maintenance of a constant rhythmic gait to keep up with the steady pace of the treadmill. This artificial walking does not necessarily translate into easier walking in hall corridors or improved ability to navigate curbs and uneven sidewalks in daily life. Meaningful clinically important differences have been defined for the 6-minute walk and are anchored in patient-reported outcomes that include mobility and quality of life. An excellent coefficient of variation for repeated 6-minute walk testing has been achieved in a PAD population and is better than that reported by Hiatt et al for treadmill testing. Thirty years of experience measuring peak treadmill walking time does not alone justify its continued use as a primary outcome. The evidence suggests that peak treadmill walking time does not optimally measure outcomes that matter most to PAD patients.
The Treadmill Is a Better Functional Test Than the 6-Minute Walk Test in Therapeutic Trials of Patients With Peripheral Artery Disease

William R. Hiatt, R. Kevin Rogers and Eric P. Brass

Circulation. 2014;130:69-78
doi: 10.1161/CIRCULATIONAHA.113.007003

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/130/1/69