

Examination of the Six Minute Walk Test to Determine Functional Capacity in People with Chronic Heart Failure: A Systematic Review

Brittany Pollentier, PT, DPT;¹ Sonya L. Irons, PT, MPT, CCS;¹ Courtney Manfredi Benedetto, PT, MPT;¹ Anne-Marie DiBenedetto, PT, MPT;¹ Dana Loton, PT, MPT;¹ Rebecca Diane Seyler, PT, DPT;¹ Mira Tych, PT, MPT;¹ Roberta A. Newton, PT, PhD, FGSA²

¹ Transitional DPT students, Temple University, Philadelphia, PA

² Professor, Department of Physical Therapy, College of Health Professions, Temple University, Philadelphia, PA

ABSTRACT

Purpose: To evaluate the reliability and validity of the six-minute walk test (6MWT) with respect to its ability to predict functional capacity in patients with chronic heart failure. **Methods:** A systematic review was performed via 8 databases to assess relevant English language full-text articles published from January 1, 1980 to October 31, 2009. Participant characteristics, interventions, reliability, validity, and predictive value for each article with respect to functional capacity as defined by peak VO_2 levels were extracted and compared. Quality Assessment of Diagnostic Accuracy Studies (QUADAS) scores were determined for each study. **Results:** Fourteen studies met the selection criteria. Comparison of the studies investigating reliability shows that the 6MWT has good reproducibility. The 6MWT demonstrates moderate correlation with peak VO_2 levels, and ability to predict VO_2 (functional capacity) dependent on distance walked. Cut-off distances vary from 300 to 490 meters depending on the study; if total distance walked remains equal or less than these values, the 6MWT retains its strong predictive value. **Conclusion:** The 6MWT has good reliability, moderate validity, and a significant ability to predict functional capacity in patients with CHF who do not walk greater than 490 meters.

Key Words: six-minute walk test, chronic heart failure, functional capacity, peak VO_2

INTRODUCTION

The incidence and prevalence of chronic heart failure (CHF) in the United States is continually rising, with an estimated 670,000 new cases diagnosed per year.¹ There are approximately 5.7 million Americans living with the disease at any given time.¹ Although advances in medicine have led to longer overall survival rates for patients diagnosed with CHF, especially following myocardial infarction and improved management of hypertension, there has also been a steady increase in mortality rates for these patients.² The projected 5-year survival rate for patients diagnosed

with heart failure is approximately 50% with 25% of patients dying within the first year.¹

Common causes of CHF include past myocardial infarction, coronary artery disease, dilated and hypertrophic cardiomyopathy, myocarditis, abnormal heart valves, congenital heart disease, hypertension, lung disease, diabetes, and less commonly, hypothyroidism, severe anemia, and/or cardiac arrhythmias.¹ Patients with CHF typically present with various nonspecific symptoms such as fatigue, lethargy, exercise intolerance, lower extremity edema, coughing, and shortness of breath.³ While symptomatology makes diagnosis difficult, objective information can be obtained from cardiac workups inclusive of symptoms, echocardiogram showing left ventricular ejection fraction, and stress test to examine exercise tolerance. Patients' symptoms and functional exercise capacity are used to classify the severity of heart failure based on the New York Heart Association (NYHA) guidelines (stages I-IV).

Management of CHF focuses around medication regimens, usually inclusive of angiotensin converting enzyme (ACE) inhibitors, diuretics, digitalis, beta blockers, calcium channel blockers, angiotensin II receptor blockers, potassium supplements, and cholesterol-lowering drugs. Surgery based upon rectifying the cause of heart failure may be indicated in some occasions. Lifestyle changes such as smoking cessation, weight-maintenance, participating in exercise, and eating a healthy diet are all options that patients are encouraged to do in order to manage CHF symptoms. Patients with heart failure generally have reduced exercise capacity, and two of the main symptoms in heart failure are exercise intolerance and generalized fatigue.⁴ This makes it difficult for these patients to participate in exercise and many times, deters them from doing so for fear of exacerbating symptoms. This leads the patient into a vicious cycle of progressive decline in function and deconditioning due to lack of exercise.

It is becoming increasingly crucial for medical professionals to use valid and reliable tests and measures in practice to determine outcomes. This is especially true in the physical therapy realm, where the move to direct access and increasingly stringent insurance guidelines cause physical therapists to require tests with high validity to ensure patient progress is well-monitored and most of all, accurate. It is necessary to have valid tests that are simple and quick to perform in the clinic in order to preserve time

Address correspondence to: Sonya Irons, PT, MPT, CCS, Madonna Rehabilitation Hospital, 5401 South Street, Lincoln, NE 68506 (sironsm@madonna.org).

management, efficiency, and to have a way to retest and monitor progress easily. While it is known that using VO_{2max} is a valid measure of functional capacity for patients with CHF,⁵ it is difficult, if not impossible, for most physical therapists to acquire this information in the clinic. The six-minute walk test (6MWT) is frequently used to determine functional capacity in patients. The 6MWT is performed on a flat surface, with a predetermined distance measured and marked off. Patients are asked to walk self-paced back and forth between the 2 marked points for 6 minutes. Rests are allowed, and administrators are able to provide standardized feedback during and at the end of the test. Patients are allowed to use an assistive device, but the tester should not walk with the patient during the test.⁶ The laps are counted, and a distance walked in 6 minutes is determined.⁶ This systematic review was formulated to examine the validity and reliability of the 6MWT in measuring functional capacity in patients with CHF.

METHODS

This systematic review was completed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISM) protocol. This protocol includes a 27-item checklist endorsed and adopted by the *Physical Therapy Journal*.⁷ The following 8 databases were searched from January 1, 1980 to October 31, 2009: CINAHL, Cochrane Central Register of Controlled Trials, Cochrane Systematic Reviews, Cochrane Methodology, EBM Reviews, Medline (via OVID), Physiotherapy Evidence Database (PEDro), and Pub Med.

Inclusion criteria for this systematic review included: Caucasian, male or female, primary diagnosis of CHF (all stages), studies that included reliability and/or validity of the 6MWT to determine functional capacity, and published in English language. Exclusion criteria for this systematic review included: postsurgical patients including heart transplant, CHF not a primary diagnosis, 6MWT not the focus of the study, pharmacological interventions as study focus, and response of exercise training as study focus. An exact list of medical subject heading (MeSH) terms were used to search the databases including pairing the following words: "heart failure," "submaximal exercise tolerance," "treadmill test," "stress test," "exercise test," "exercise tolerance test," "exercise capacity," "6 minute walk test," "six minute walk test," "prognosis," "prognostic indicator," "function," "functional capacity," "functional outcomes," "reliable," "reliability," "validity," "valid," "outcome measure," "reproducibility," "reproducibility of results," and "reproducibility of findings."

To ensure a thorough search, the 8 databases were divided among groups of 2 authors each. Each reviewer in each group independently searched each database and recorded results. Initial searches using MeSH terms resulted in 16,954 potential studies. The titles, abstracts, and/or full text articles for these search results were initially independently screened by 2 reviewers per database to determine their appropriateness in this systematic review. Studies were eliminated if they did not meet the exclusion and/or inclusion criteria. Secondary or hand searches by one reviewer

for each database were also conducted. After eliminating studies because of the aforementioned reasons, applicable studies were then further eliminated if they were duplicate studies (for example, the same study was found in multiple databases). The pool of applicable studies was then narrowed to 109 potential studies from all databases searched. These studies were then further analyzed for their appropriateness by 2 independent reviewers each. Ninety-five studies were then eliminated because they did not abide by the set inclusion or exclusion criteria, or they were duplicates. This process ultimately resulted in 14 applicable studies for this systematic review (Figure 1).

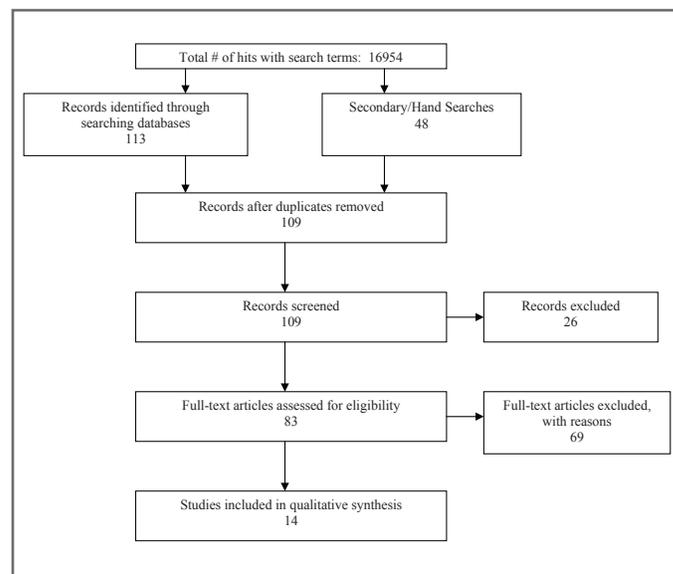


Figure 1. Details of flow of information. Adapted from Moher.⁷

To assess the quality of the 14 studies, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) scale was used.⁸ The QUADAS scale was selected due to its ability to assess the quality of diagnostic accuracy studies.⁸ This scale has 14 questions, which includes questions related to bias, variability, and some quality of reporting (Appendix 1).⁸ Two reviewers who were blinded to the scores of the other reviewers assessed a QUADAS score for each study. An arbitrary QUADAS score of at least 50% was considered eligible for inclusion in this study. Inter-rater reliability (ICC = 0.889) was calculated using SPSS statistical program to determine the agreement between raters when using the QUADAS tool on one of the 14 studies included in this systematic review. If differences occurred between reviewers for QUADAS score, then they further discussed and reached a consensus regarding the appropriate score.

For data extraction from the 14 selected articles, 2 reviewers independently extracted data related to population, interventions, reliability, validity, reproducibility, predictive value, and functional capacity as related to peak VO_2 (Table 1). Functional capacity has been defined as the ability of the individual to perform aerobic work as defined by maximal oxygen uptake (VO_{2max}) and therefore the ability for patients to perform day-to-day activities.^{9,10}

Table 1. Study Characteristics, QUADAS Score, and Results of Studies

Source	Population	Tests Performed	QUADAS Score	Results
Cahalin ¹⁴ 1996	Patients with CHF; mean LVEF of 20% ± 6%	Pulmonary function testing, symptom-limited bicycle ergometry exercise testing with radionuclide angiography and cardiopulmonary response, 6MWT with cardiopulmonary response (repeated for 20 patients) by blinded physical therapists, cardiac catheterization	13 of 14	<p>Reproducibility:</p> <ul style="list-style-type: none"> - ICC = 0.96 for 6 MWT <p>Validity:</p> <ul style="list-style-type: none"> - 6 MWT moderately correlated with peak VO₂ (r = 0.64, p = 0.0001) <p>Predictive Value:</p> <ul style="list-style-type: none"> - 6MWT a significant predictor (upon multivariate analysis) of peak VO₂ (r=0.64, p<0.0001)
Green ¹⁷ 2001	Patients from cardiac transplant unit with CHF (NYHA classes II-IV); LVEF of < 35%; mean age 56.7 +/-1.5 years	1 st experiment: SWT administered twice 2 nd experiment: SWT, 6 MWT, treadmill test for VO ₂ max	11 of 14	<p>Validity:</p> <ul style="list-style-type: none"> - Correlation of peak VO₂ and distance ambulated during 6 MWT was moderate but not statistically significant (r = 0.67) - Correlation of peak VO₂ during 6MWT and peak VO₂ during treadmill test was moderate but not statistically significant (r = 0.70)
Guyatt ¹⁰ 1985	Patients with CHF (NYHA class II-IV); also included a respiratory group; mean age 64.7+/-8.3 years	Maximal exercise bicycle ergometer test, 6 MWT was repeated 6 times per patient (2 weeks between tests)	7 of 14	<p>Reproducibility:</p> <ul style="list-style-type: none"> - Article reports reproducibility, but no statistics given <p>Validity:</p> <ul style="list-style-type: none"> - 6MWT poorly correlated with maximal cycle ergometer (r=0.42, p<0.001)
Guyatt ¹⁵ 1985	Patients with CHF (NYHA class II-IV); mean age 64.7+/-8.3 years	6 MWT and cycle ergometry performed 6 times, 2 weeks in between tests	10 of 14	<p>Reproducibility:</p> <ul style="list-style-type: none"> - ICC for 6 MWT for all visits (6 visits total) = 0.921 <p>Validity:</p> <ul style="list-style-type: none"> - 6 MWT moderately correlated with maximal cycle ergometer exercise test (r= 0.579, p=0.001)
Lipkin ¹⁸ 1986	Patients with congestive heart failure (NYHA classes II and III); mean age 58 years (range 36-68)	Maximal treadmill exercise test with pulmonary testing, 6 MWT	10 of 14	<p>Validity:</p> <ul style="list-style-type: none"> - Curvilinear relationship existed between max VO₂ and distance walked in 6 MWT (no data given)
Lucas ¹¹ 1999	Patients with heart failure having LVEF<35%, mean age 52 +/-17 years	6 MWT and bicycle exercise test performed 2 hours apart with VO ₂ max testing	12 of 14	<p>Validity:</p> <ul style="list-style-type: none"> - 6 MWT poorly correlated with peak VO₂ (10-20mL/kg per 6MWT moderately correlated with peak VO₂ (>20mL/kg or <10mL/kg per minute) bicycle testing (r = .57)
Morales ¹⁶ 1999	Patients with CHF symptoms for at least 6 months, NYHA class II-IV; LVEF of 23% ± 8%	2 to 3 of 6MWT, SWT and symptom-limited bicycle ergometry test within 2 weeks of each other	11 of 14	<p>Reproducibility:</p> <ul style="list-style-type: none"> - 6MWT with good reproducibility only after a first trial (i.e. no significant differences between 2nd & 3rd trials (r=0.98, p=0.33)) <p>Validity:</p> <ul style="list-style-type: none"> - 6MWT moderately correlated with peak VO₂ (r=0.69, p<0.001) <p>Predictive value:</p> <ul style="list-style-type: none"> - 6 MWT considered a significant predictor of peak VO₂ < 14 mL/kg/min (ROC curve analysis=0.83, p = 0.02), but upon multi- variate analysis not considered to be an independent predictor - Best distance cut-off to predict peak VO₂ (<14 mL/kg/min) with 6MWT = 450 m (sensitivity 80%, specificity 83%)
Myers ¹⁹ 2006	Male patients with stable heart failure; EF < 40%; mean age 68+/-12 years	Symptom-limited incremental treadmill exercise test, pulmonary function tests, 6 MWT	9 of 14	<p>Validity:</p> <ul style="list-style-type: none"> - 6MWT modestly correlated with peak VO₂ r= 0.49, p<0.01) <p>Predictive value:</p> <ul style="list-style-type: none"> - 6MWT not a significant predictor (upon multivariate analysis) of peak VO₂ (no data given)
O'Keefe ²³ 1998	Patients with heart failure (NYHA classes I-IV), mean age 82 years (range 74 to 92)	6 MWT repeated twice	8 of 14	<p>Reproducibility:</p> <ul style="list-style-type: none"> - ICC for 6 MWT = 0.91
Peeters ²⁰ 1996	Patients with heart failure (NYHA classes II and III), mean age 81.3 (SD 5.6 years) and untrained, healthy, controls	6MWT and progressive, symptom-limited, treadmill test.	10 of 14	<p>Distanced walked during the 6MWT correlated with the distance walked on the treadmill test (r=.688; p<.001).</p> <p>Significant lower symptomatic VO₂ max in NYHA class III than in NYHA class II (p=.032).</p>

Source	Population	Tests Performed	QUADAS Score	Results
Pulz ²¹ 2008	Patients with CHF (NYHA class II-IV); LVEF < 35%; mean age 51.3 +/- 10.2 years	Maximal incremental treadmill exercise test to determine VO ₂ max, 2 6MWTs 30 minutes apart per patient administered by physical therapist, 2 incremental shuttle walk tests	11 of 14	Reproducibility: - mean bias between repeated 6MWTs close to zero +/-1.96 SD, 95% CI 7+/- 40 meters Validity: - 6MWT correlated well with pVO ₂ (r= 0.76, p < 0.001) Predictive value: - 6MWT a significant predictor (upon ROC curve analysis = 0.91) of peak VO ₂ - Best distance cut-off to predict pVO ₂ with the 6 MWT = 490 m (sensitivity 83%, specificity 83%)
Riley ²² 1992	Patients with CHF (NYHA classes II-IV); mean LVEF of 26%; mean age 65.2 years (range 48-76)	Symptom-limited max treadmill test to determine VO ₂ max, 3 6MWTs with VO ₂ max taken	10 of 14	Reproducibility: - 6MWT with good reproducibility (coefficient of variation = 6.71) Validity: - 6MWT distance correlated well with peak VO ₂ as measured during treadmill test (r=0.88, P < 0.001)
Roul ¹² 1998	Patients with heart failure (NYHA classes II or III); mean age 59+/-11 years	Symptom-limited bicycle ergometry exercise test with VO ₂ max testing, 6 MWT performed, repeated twice on the same day in 40 subjects	9 of 14	Reproducibility: - ICC for 6 MWT = 0.82 Validity: - 6MWT moderately correlated with peak VO ₂ only if distance walked was <= 300 m (r=.65, p=0.011) - 6MWT a significant predictor of peak VO ₂ but only with cut-off distance of 300 m (ROC curve = 0.933; sensitivity 89%, specificity = 60%, PPV = 80%)
Zugck ¹³ 2000	Patients with congestive heart failure (NYHA classes I-III); LVEF < 35%; mean age 54+/-12 years	Symptom-limited bicycle exercise test performed; 6 MWT performed by patients with supervision of RN blinded to exercise test results	13 of 14	Reproducibility: - ICC for 6MWT = .96 Validity: - 6MWT moderately correlated with peak VO ₂ (r = .689 in men, r = .59 in women, p = 0.001) Predictive value: - 6MWT moderately correlated with peak VO ₂ over time: - After 263 days +/-14: r=0.71, p<0.001 - After 381 days +/- 170: r=0.74, p<0.001

Abbreviations: LVEF = Left Ventricular Ejection Fraction; EF = Ejection Fraction

RESULTS

Fourteen articles were deemed relevant for this review. The primary purpose of the majority of these studies was to investigate the utility of the 6MWT for patients with CHF. Samples varied from 14 to 63 with 3 studies reporting sample sizes larger than 100 participants.¹¹⁻¹³ Two of the studies, Cahalin¹⁴ and Zugck,¹³ used testers who were blinded to the results of the studies.

A common factor across many of the studies was the use of maximal exercise testing to acquire information regarding VO₂max. Seven of the studies used symptom-limited bicycle ergometry,¹⁰⁻¹⁶ because bicycle ergometry was less troublesome for the patient to perform, especially in the elderly population. Six studies used maximal symptom-limited treadmill testing.¹⁷⁻²² The shuttle-walk test was used in several studies as a comparison against the 6MWT,^{16,17,21} one of which showed high predictive value for VO₂max.²¹ Seven studies examined test-retest reliability; one of which showed good reliability and reproducibility.^{12-16,22,23} One study²³ focused solely on reliability of the 6MWT, finding significant reliability in patients with no change in cardiac status upon follow-up (ICC=0.91), as well as in all patients across the entire sample (ICC > 0.75).

Two studies found that 6MWD ≤ 300 m was significantly able to predict likelihood of death or poorer outcomes in 6 months (p = .04)¹⁴ and worse prognostic outcomes (p

= .013).¹² Two other studies determined that distances ≤ 450m (r = 0.98, p = 0.33) and ≤ 490m (r = 0.76, p < 0.001), respectively also correlated with VO₂max.^{16,21} Peeters et al²⁰ determined that the distance covered in a 6MWT correlated to distance walked on a treadmill (p < .001). Another 4 studies showed high, but not statistically significant correlations between VO₂max and the 6MWD.^{13,18,19,22} Lipkin¹⁸ also determined that patients with high VO₂max had lower variations in 6MWD, conversely, patients with low VO₂max (< 14 ml/kg/min) had higher variability of 6MWD. Four of the studies included did not find a significant correlation between 6MWD and VO₂max from exercise testing.^{10,11,15,17}

Synopsis of Each Study and Author Recommendations Cahalin¹⁴

In order to evaluate the utility of the 6MWT in assessing patients with CHF, 45 patients with symptomatic CHF underwent symptom limited cardiopulmonary exercise testing and the 6MWT. The 6MWT correlated moderately with peak VO₂ (r = .64, p < 0.001) in patients with heart failure referred for heart transplantation. Distance ambulated less than 300 m was predictive of hospitalization or death in the next 6 months (p = 0.04). The 6MWT was inferior to peak VO₂ and %VO₂ for prediction of long-term survival; therefore, its use as a test of long-term survival is uncertain.

The 6MWT may predict maximal bicycle exercise testing in patients with advanced CHF. From this study, equations were formulated that predict peak VO_2 from the distance walked with 6MWT. Lastly, the 6MWT predicted peak VO_2 and short-term event free survival of patients with advanced heart failure.

Green¹⁷

This study of 7 subjects examined the 10 m shuttle walk test (SWT) compared with the 6MWT and a maximal symptom-limited treadmill test in patients with CHF. Heart rate, rate of perceived exertion, and ambulatory VO_2 were examined. Distance ambulated during the 6MWT was not significantly correlated to predict VO_2 or functional capacity. The 6MWT provides an assessment of the patient's ability to perform submaximal activities of daily living but does not provide a consistent prediction of mortality in patients with CHF. Green and colleagues suggested that the SWT is a more predictive measure of peak VO_2 than 6MWT. Peak HR, RPE were similar from the SWT and the treadmill peak VO_2 test compared with VO_2 during 6MWT which was lower, suggesting the SWT elicits a more comparable exercise response and thus is a more valid measure of functional capacity than the 6MWT in patients with CHF. The physiologic responses observed during the 6MWT were not graded as in the SWT and the peak VO_2 test, suggesting the 6MWT does not elicit the same physiologic response and is not as predictive as previously thought in testing functional capacity.

Guyatt¹⁰

In order to assess the potential value of the 6MWT as an objective measure of exercise capacity in patients with CHF, the test was administered 6 times over 12 weeks to 18 patients with CHF, who also underwent cycle ergometer testing and conventional functional testing. During the 6MWT, 9 patients received encouragement and 9 did not. Results of the 6MWT were consistent after the first 2 trials ($p < .001$). Encouragement influenced the scores ($p < .02$); therefore, the 6MWT should be standardized, eg, without encouragement. The low correlation of 6MWT scores with the cycle ergometer test ($r = .42$) suggest the 6MWT measures more functional activities as opposed to laboratory exercise capacity. The authors concluded that the 6MWT is reproducible but has low correlation with maximal exercise capacity for patients with CHF.

Guyatt¹⁵

In order to clarify the usefulness of available measures of functional status and exercise capacity, 43 patients with lung disease and CHF were given 4 questionnaires, the 6MWT, and a bicycle exercise test. Correlations between the questionnaires and the 6MWT were $r = .0473$ -.0590 but no correlation was found between the cycle testing and the questionnaires. Correlation between the 6MWT and bicycle test was moderate ($r = .579$). Reproducibility testing showed good results with the 6MWT. Correlation of the 6MWT with the questionnaires and with cycle ergometer indicates that the 6MWT is a good measure of how

the patients feel, how they function when performing daily activities, and of their exercise capacity as measured in the laboratory. The 6MWT was considered superior to the questionnaires in that it is objective and reproducible. The 6MWT appears to measure the patient's ability to undertake physically demanding activities of daily living and may be a useful tool to measure cardiorespiratory response after intervention.

Lipkin¹⁸

Exercise capacity of 26 patients with NYHA class II-III and 10 normal subjects with a mean age of 58 (range 36-68) yrs old was assessed by determining the oxygen consumption attained during a maximal exercise test and by measuring the distance walked during the 6MWT. There were significant differences in maximal oxygen consumption between normal subjects, patients with class II heart failure and class III heart failure ($p < 0.001$). The distances walked in the third trial of the 6MWT were 683 meters in the normal subjects, 559 meters in class II patients, and 402 meters in class III patients. Maximal oxygen consumption was related in a curvilinear manner to the distance walked during the 6MWT. The findings of the study indicate that the 6MWT can objectively assess exercise capacity in patients with CHF. The patients preferred the 6MWT to the treadmill, finding it easier and more closely related to their daily physical activity. This finding may explain the wide range of distances covered in 6 minutes among those patients with low maximal oxygen consumption. Some may have been poor performers on the treadmill test and gave a better demonstration of their abilities during the 6MWT. The authors conclude that in patients with CHF the 6MWT provides an objective assessment of exercise capacity that could usefully supplement clinical information obtained from the history and physical examination.

Lucas¹¹

The purpose of the study was to investigate whether the 6MWT distance was related to peak VO_2 for individuals with advanced heart failure. A total of 307 subjects (mean age of 52 yrs, SD 13 yrs) and an ejection fraction less than 35% were studied. For patients with a peak VO_2 of 10-20 mL/kg/min, the 6MWT was poorly correlated with peak VO_2 ($r = .28$). For patients with higher (>20 mL/kg/min) or lower (<10 mL/kg/min) peak VO_2 , 6MWT moderately correlated with peak VO_2 ($r = .57$). The 6MWT distance was not well correlated with peak VO_2 and therefore considered not valid to predict peak VO_2 for patients with advanced heart failure.

Morales¹⁶

In this study, investigators assessed the 6MWT and the Shuttle Walk Test (SWT) with respect to peak VO_2 , as measured by cycle ergometry, to determine the tests' utility as a measure of functional capacity in patients with CHF. Similar to Riley,²² the 6MWT had good reproducibility only after the first of 3 trials, with a significant difference in distance walked between trial 1 and 2 ($p = 0.03$), but not between trials 2 and 3 ($r = 0.98$, $p = 0.33$). The 6MWT was moder-

ately correlated with peak VO_2 levels ($r = 0.69$, $p < 0.001$). It was considered a significant predictor of peak $\text{VO}_2 < 14$ mL/kg/min, with 83% accuracy ($p = 0.02$), with the best cut-off distance being 450 meters (sensitivity 80%, specificity 83%). For distances less than 450 meters, the 6MWT lost its significance as a peak VO_2 predictor. Findings of a multivariate analysis indicated that only SWT distance ($p < 0.001$) and not 6MWT distance was considered to be an independent predictor of peak VO_2 . The SWT was also only reproducible after a practice test, but overall, it was correlated more strongly with peak VO_2 than the 6MWT.

The authors suggest that one reason for the stronger correlation and predictive value of the SWT was because both the cycle ergometry test and the SWT are symptom limited, unlike the 6MWT which is time limited. Analysis of the clinical data also suggested that the SWT was a better correlate to the exercise test in that the SWT was significantly more strenuous for the patients than the 6MWT [with respect to maximum heart rate, systolic blood pressure, and rate of perceived exertion (RPE)], but was not significantly different when compared to the cycle ergometry test (with respect to systolic blood pressure and RPE). They concluded that the 6MWT appears inferior to the SWT with respect to predicting functional capacity as defined as peak VO_2 , although still considered to have good reproducibility, moderate correlation with peak VO_2 levels, and is 83% accurate in predicting peak $\text{VO}_2 (<14$ mL/kg/min) as long as distance walked is greater than 450 meters.

Myers¹⁹

Forty-one patients with CHF were assessed with commonly used function, health status, and quality of life measures, including maximal cardiopulmonary exercise testing, the Duke Activity Status Index (DASI), the Veterans Specific Activity Questionnaire (VSAQ), the Kansas City Cardiomyopathy Questionnaire (KCCQ), and the 6MWT. Peak VO_2 demonstrated a significant but low relationship with the VSAQ score ($r = 0.37$) and no relationship with the DASI. Among the KCCQ domains, peak VO_2 was significantly associated with quality of life score ($r = 0.46$) but not with physical limitations, symptoms, or clinical summary scores. The 6MWT distance was significantly though modestly related to peak VO_2 ($r = 0.49$), KCCQ physical limitation ($r = 0.53$), and clinical summary scores ($r = 0.44$). The authors suggest that commonly used functional measures, estimates of quality of life, and symptom tools in patients with CHF have modest associations with peak VO_2 and with one another. They further recommend that each test or instrument target a specific aspect of clinical status; they should not be considered interchangeable; and, that none of these measures alone is a reliable replacement for peak VO_2 .

O'Keefe²³

While this longitudinal study did not examine peak VO_2 levels as it relates to the 6MWT, it did investigate the reproducibility and responsiveness to change of the 6MWT, as well as the chronic heart failure questionnaire (CHQ), in more elderly patients (mean age 82 yrs, range 74-92) with heart failure. The 6MWT demonstrated excellent re-

producibility (after re-testing 3 and 8 weeks later) with an intraclass correlation coefficient $R = 0.91$. Data were only analyzed for the 24 of 45 patients who performed the second test. They reported no major overall change in cardiac status; those who reported that they had gotten better or worse were excluded from the second test. This global rating of change had good to moderate correlation with the dyspnea quality of life (QOL) ($r = .70$), fatigue QOL ($r = .69$), and total QOL scores ($r = .77$) as well as with walk distance ($r = .78$), but less moderate correlation to emotional QOL scores ($r = .46$) and heart failure score (CHQ) ($r = -.56$). Since questionnaires were involved in this study, the subjective nature of patient's answers could present bias. Despite possibly questionable assurance of lack of change with patient functional status between tests, the 6MWT demonstrated adequate reproducibility and responsiveness in elderly adults with CHF.

Peeters²⁰

In this prospective trial, 37 patients were included on a 6MWT and a symptom-limited treadmill test. Participants with CHF NYHA classes II (16 patients) and III (10 patients) were compared with healthy, untrained controls (10 patients). Interestingly, all patients preferred the 6MWT to the treadmill test. This study had 8 participants (22%) drop out due to being unable to perform the treadmill test. The 6MWT distance was significantly different between NYHA classes II and III ($p < .001$), and control and NYHA class III ($p < .001$). No significant difference between control and NYHA class II. Significant lower symptomatic VO_2 max occurred in NYHA class III than in NYHA class II ($p = .032$). The authors concluded that the 6MWT was correlated with treadmill distance, well tolerated by the elderly, and provided functional differentiation between NYHA classes II and III.

Pulz²¹

This study investigated the incremental shuttle walk test (ISWT) and the 6MWT with respect to test-retest reproducibility and accuracy in predicting peak VO_2 in patients with stable CHF. The investigators used maximal incremental cardiopulmonary exercise testing to obtain peak VO_2 levels. All 63 patients underwent cardiopulmonary exercise testing and then on a separate day were randomly assigned to 4 walk test groups (2 6MWT, 2 ISWT) with the repeated test performed 30 minutes later. Patients were verbally encouraged throughout the 6MWT each minute. With respect to test reproducibility, there was no significant difference in mean values for the 6MWT, with the mean bias close to 0 (1.96 SD). The 95% CI was 40 meters per Bland-Altman analysis. The 6MWT correlated well with peak VO_2 levels ($r = 0.76$, $p < 0.001$), although there was wide variability in distance walked for the given peak VO_2 level of 14 mL/kg/min. The 6MWT was a significant predictor of significantly reduced peak VO_2 (14 mL/kg/min or lower) with ROC curve = 0.91. The best predictive cut-off distance was 490 meters (sensitivity 83%, specificity 83%). A significant limitation with this study with respect to predicting functional capacity as measured by peak VO_2 was that measurements were

taken over a short span of time. The patients were also relatively young for people with CHF (51.3 yrs SD 10.2 yrs) and had a relatively “well-preserved” functional capacity, which may limit the predictive value of these tests. The results of the study suggest that the ISWT and the 6MWT have comparably modestly good reproducibility and accuracy in predicting significantly reduced peak VO_2 in patients with stable CHF.

Riley²²

In this study investigators assessed the reproducibility of the 6MWT and its validity as a measure of functional capacity in patients with CHF. VO_2 was measured not only during a treadmill exercise test, but also during performance of the 6MWT, providing for additional comparison. The 6MWT had good reproducibility after the first of 3 trials with a significant difference in distance walked between trial 1 and 2, but not 2 and 3; the 6MWT comparing trial 2 and 3 had a co-efficient of variance of 6.71. The 6MWT distance had good correlation with treadmill-measured peak VO_2 ($r = 0.88$, $p < 0.0001$), as did the VO_2 measured during the 6MWT ($r = 0.90$, $p < 0.0001$). Given that the VO_2 measured during the 6MWT was less than the exercise treadmill test peak VO_2 for all patients, it provides support for the notion that the 6MWT is a submaximal exercise test. Therefore the 6MWT provides for less variability in gait dynamics among subjects and a more accurate assessment of functional capacity. Patients may need to perform a practice trial of the 6MWT as reproducibility appears to only be accurate after the second test. The 6MWT has good correlation with peak VO_2 , but is likely a better measure of submaximal exercise.

Roul¹²

The purpose of this study was to evaluate the potential of the 6MWT compared with peak VO_2 in predicting outcome of patients with New York Heart Association (NYHA) class II or III heart failure. A total of 121 male and female subjects between the ages of 18-76 (mean age 59 yrs SD 11 yrs) with diagnosed heart failure for at least 3 months were included. Reliability was established for 40 of the 121 subjects by repeating the 6MWT twice in the same day ($\text{ICC} = 0.82$). This study has detection bias in that the 6MWT was not repeated in the whole study population. And although the intraclass correlation was high, it cannot be concluded that any improvement would not be observed, particularly in those patients walking shorter distances, which could modify to some extent the results reported. When using a threshold of 300 m the sensitivity of the 6MWT was 89%, specificity 60%, and positive predictive value 80%. There was a moderate correlation between distance walked and peak VO_2 when a distance of 300 m was used as a threshold ($r = 0.65$).

Zugck¹³

The purpose of this study was to prospectively examine whether the distance ambulated during a 6MWT correlated with peak VO_2 . A total of 113 subjects with stable congestive heart failure (NYHA I-III) and ejection fraction less

than 35% underwent symptom-limited bicycle cardiopulmonary exercise test while measuring peak VO_2 , and performed the 6MWT. Reliability was established by repeating the 6MWT for 10 subjects on 3 consecutive days ($\text{ICC} = .96$). There was a moderate correlation between distance ambulated and peak VO_2 ($r = .69$ in men, $r = .59$ in women, $p = 0.001$). The 6MWT has predictive value of determining peak VO_2 over time exhibited by moderate correlations between 6MWT and peak VO_2 after 263 (SD 14) days ($r = 0.71$, $p < 0.001$) and after 381 (SD 170) days ($r = 0.74$, $p < 0.001$). The 6MWT correlated closely with peak oxygen uptake, has the ability to predict individual peak oxygen uptake, and can provide prognostic information of peak oxygen uptake.

DISCUSSION

Attempts to measure functional capacity in persons with chronic heart failure have varied from questionnaires, grouping into NYHA classes, peak VO_2 measurement, and submaximal exercise testing. The 6MWT is a low cost, simple test that requires little equipment and shows good to excellent test-retest reliability across the literature. The results of the test, however, are not reliable until after the first trial;^{16,22} this will decrease the learning effect of the test.

The Shuttle Walk Test (SWT) may better correlate to a standard cardiopulmonary symptom-limited incremental exercise test than the 6MWT, as the 6MWT is distance-limited, self-paced, less strenuous, and therefore more suggestive of submaximal functional capacity.¹⁷ Pulz²¹ does not agree with Green¹⁷ with respect to the relationship of the SWT and 6MWT with peak VO_2 , perhaps due to the larger sample size of Pulz's²¹ study.

Limitations of the studies relevant to this systematic review include a lack of consistency for administration of the 6MWT, including whether encouragement was given and the amount, whether the administrator of the test had experience in giving the test and the reliability of the tester, and whether the tester was blinded to the other tests and previous 6MWT results. When considering the accuracy of the peak VO_2 measurements, many of the articles used cycle ergometry for their measurements, which may give results of peak VO_2 as much as 10% less than maximal treadmill exercise test peak VO_2 .¹⁶ Many studies used a population of only certain NYHA classes of CHF instead of a combination of all of the classes, which would decrease the ability to use the results in those classes of CHF that weren't included. Also, the different terminology and definitions of functional capacity in the literature made it difficult to extrapolate this information. By limiting our article selection to the English language only, results may have bias through incomplete retrieval.

Future studies should focus on increasing the quality of the research. The items listed above should be better established, such as clearly defined 6MWT administration, how much encouragement is given, how much experience the testers have, blinding of testers, use of a control group, use of the maximal treadmill exercise test for a measure of peak VO_2 , and inclusion of patients from each of the four NYHA classes of CHF.

CONCLUSION

This systematic review focused on the reliability and validity of the 6MWT for assessing functional capacity in people with CHF. The correlation of the 6MWT to peak VO_2 is moderate to good, with 6MWT being between 83% and 91% accurate in predicting peak VO_2 if distance walked is less than 450 to 490 meters, respectively.^{16,21} Stronger correlations are observed when individuals with advanced heart failure walk a distance less than 300 meters in 6 minutes or have a low (<10 mL/kg/min) peak VO_2 .¹² The 6MWT is less of a predictor of peak VO_2 in patients who have higher peak VO_2 since the test is time limited and not symptom limited, making VO_2 max difficult to predict in more fit patients.¹⁶ The 6MWT may give a better representation of a patient's ability to perform submaximal activities of daily living.^{17,22} Using the 6MWT as a prognostic indicator may also be useful in serial evaluation of the patient's status and/or response to therapeutic interventions. In conclusion, the 6MWT has moderate ability to predict peak VO_2 and functional capacity in those patients with CHF who walk less than 490 meters.

Appendix A. Quality Assessment of Diagnostic Accuracy Studies: QUADAS⁸

1. Was the spectrum of patients representative of the patients who will receive the test in practice?
2. Were selection criteria clearly described?
3. Is the reference standard likely to correctly classify the target condition?
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of the index test result?
7. Was the reference standard independent of the index test (ie, the index test did not form part of the reference standard)?
8. Was the execution of the index test described in sufficient detail to permit replication of the test?
9. Was the execution of the reference standard described in sufficient detail to permit its replication?
10. Were the index test results interpreted without knowledge of the results of the reference standard?
11. Were the reference standard results interpreted without knowledge of the results of the index test?
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
13. Were uninterpretable/intermediate test results reported?
14. Were withdrawals from the study explained?

REFERENCES

1. American Heart Association. 10/27/09. Available at: <http://www.americanheart.org>. Accessed November 13, 2009.
2. Kostis J, Davis B, Cutler J, et al. Prevention of heart failure; antihypertensive drug treatment in older persons with isolated systemic hypertension. SHEP Cooperative Research Group. *JAMA*. 1997;278(3):212-216.
3. Ingle L, Cleland J, Clark A. Perception of symptoms in out of proportion to cardiac pathology in patients with "diastolic heart failure." *Heart*. 2008;94(6):748-753.
4. Jonsdottir S, Andersen K, Sigurosson A, Sigurosson S. The effect of physical training in chronic heart failure. *Eur J Heart Fail*. 2006;8(1):97-101.
5. Bartlo P. Evidence-based application of aerobic and resistance training in patients with congestive heart failure. *J Cardiopulm Rehabil*. 2007;27(6):368-375.
6. Enright P, McBurnie M, Bittner V, et al. The 6-minute walk test: a quick measure of functional status in elderly adults. *Chest*. 2003;123:387-398.
7. Moher D, Liberati A, Tetzlaff J, Altman G, Group P. Reprint—preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Phys Ther*. 2009;89(9):873-880.
8. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3(25):1-25.
9. Arena R, Myers J, Williams M, et al. Assessment of functional capacity in clinical and research settings: a scientific statement from the American Heart Association Committee on Exercise, Rehabilitation, and Prevention of the Council on Clinical Cardiology and the Council on Cardiovascular Nursing. *Circulation*. 2007;116(3):329-343.
10. Guyatt G, Sullivan M, Thompson P, et al. The six-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J*. 1985;132:919-923.
11. Lucas C, Stevenson L, Johnson W, et al. The 6-min walk and peak oxygen consumption in advanced heart failure: aerobic capacity and survival. *Am Heart J*. 1999;138:618-624.
12. Roul G, Germain P, Bareiss P. Does the 6-minute walk test predict the prognosis in patients with NYHA class II or III chronic heart failure? *Am Heart J*. 1998;136(3):449-457.
13. Zugck C, Kruger C, Durr S, et al. Is the 6-minute walk test a reliable substitute for peak oxygen uptake in patients with dilated cardiomyopathy? *Eur Heart J*. 2000;21(7):540-549.
14. Cahalin L, Mathier M, Semigran M, Dec G, DiSalvo T. The six-minute walk test predicts peak oxygen uptake and survival in patients with advanced heart failure. *Chest*. 1996;110(2):325-332.
15. Guyatt G, Thompson P, Berman L, et al. How should we measure function in patients with chronic heart and lung disease? *J Chronic Dis*. 1985;38(6):517-524.

16. Morales F, Martinez A, Mendez M, et al. A shuttle walk test for assessment of functional capacity in chronic heart failure. *Am Heart J*. 1999;138(2 Pt 1):291-298.
17. Green D, Watts K, Rankin S, Wong P, O'Driscoll J. A comparison of the shuttle and 6 minute walking tests with measured peak oxygen consumption in patients with heart failure. *J Sci Med Sport*. 2001;4(3):292-300.
18. Lipkin D, Scriven A, Crake T, Poole-Wilson P. Six-minute walking test for assessing exercise capacity in chronic heart failure. *Br Med J*. 1986;292(6521):653-655.
19. Myers J, Zaheer N, Quaglietti S, Madhavan R, Froelicher V, Heidenreich P. Association of functional and health status measures in heart failure. *J Card Fail*. 2006;12(6):439-435.
20. Peeters P, Mets T. The 6-minute walk as an appropriate exercise test in elderly patients with chronic heart failure. *J Gerontol A Biol Sci Med*. 1996;51(4):M147-151.
21. Pulz C, Diniz R, Alves A, et al. Incremental shuttle and six-minute walking tests in the assessment of functional capacity in chronic heart failure. *Can J Cardiol*. 2008;24(2):131-135.
22. Riley M, McParland J, Standford C, Nicholls D. Oxygen consumption during corridor walk testing in chronic cardiac failure. *Eur Heart J*. 1992;13:789-793.
23. O'Keefe S, Lye M, Donnellan C, Carmichael D. Reproducibility and responsiveness of quality of life assessment and six minute walk test in elderly heart failure patients. *Heart*. 1998;80:377-382.