Systematic Review Article

Title: A systematic review of the six-minute walk test in outpatient cardiac rehabilitation: validity, reliability, and responsiveness.

Authors: R. Nicole Bellet, BPhty, MPH[1 2 3], Lewis Adams BSc PhD [2 3], Norman R. Morris BAppSc Physio, PhD [2 3].

[1] Physiotherapy Department, The Prince Charles Hospital, Metro North Health Service District, Brisbane, Queensland, Australia
[2] School of Physiotherapy and Exercise Science, Griffith University, Gold Coast, Australia
[3] Griffith Health Institute, Griffith University, Gold Coast , Australia

Key words: six-minute walk test, cardiac rehabilitation, reliability, validity, responsiveness

Word Count: **250 words (Structured Abstract)**
3048 words (Text)

References: **36**

Figures: **2**

Box: **2**

Tables: **4**

Source(s) of support: Grant from Queensland Health Community Rehabilitation Workforce Project: Community Rehabilitation Research Scheme: letter dated 24/6/07.

Acknowledgements: We would like to acknowledge Leanne Bisset MPhty PhD for her expert advice in the review of the manuscript and Ian Yang MBBS (Hons), PhD, FRACP, Grad Dip Clin Epid for his expert advice and assistance with the meta-analysis.

Competing interests: No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the authors or on any organization with which the authors are associated.

Title Page

Systematic Review Article

Title: A systematic review of the six-minute walk test in outpatient cardiac rehabilitation: validity, reliability, and responsiveness.

Key words: six-minute walk test, cardiac rehabilitation, reliability, validity, responsiveness

Word Count: 250 words (Structured Abstract)
3048 words (Text)

References: 36

Figures: 2

Box: 2

Tables: 4

Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

Structured Abstract (250 words)

Background: The six-minute-walk-test (6MWT) is a common outcome measurement in cardiac rehabilitation (CR); however, a search of the literature found no guidelines established for the use of the 6MWT in CR.

Objectives: Systematically review the validity, reliability, and responsiveness of the 6MWT in CR.

Data Sources: OvidMEDLINE, SPORTdiscus, EMBASE, CINAHL, Cochrane Reviews, Cochrane Clinical Trials between January 1948 and April 2011.

Eligibility Criteria: Studies using 6MWTs in subjects with coronary artery disease undergoing outpatient-CR were included. Non-English-language articles were excluded.

Study appraisal and methods: Quantitative and qualitative analyses were conducted including, methodology quality assessment, meta-analysis and assessment against level of evidence criteria.

Results: Fifteen articles met inclusion criteria. One high quality study was identified for reliability, six for validity and eleven for responsiveness. Meta-analysis indicated strong evidence that the 6MWT was responsive to change in clinical status following CR with an estimated mean difference (95% CI, p-value) in 6MWD of 60.43m (54.57, 66.30, <0.00001). Qualitative analysis indicated moderate evidence for the repeatability of the 6MWT in the CR population, for a 2-8% learning effect between repeated-6MWTs, for a relationship between peak heart rate during 6MWT and during cycle-exercise at ventilatory threshold and for a moderate to high correlation between the 6MWD and maximum-metabolic-equivalents achieved on symptom-limited-exercise-test.

Limitations: Few studies assessed similar aspects of validity for the 6MWT

Conclusion: Strong evidence suggests that the 6MWT is responsive to clinical change during CR outcome measurement. Intra- and inter-tester reliability of the 6MWT and its validity in the CR population requires more research.

Background

Measurement of oxygen consumption, during cardiopulmonary exercise testing, represents the gold standard for determining baseline functional capacity, training intensity, cardiovascular risk, and for the evaluation of training outcomes in the cardiac rehabilitation (CR) population(1). However, the six-minute-walk-test (6MWT), is frequently recommended (2) (3) (4) to estimate functional exercise capacity in CR rather than subjecting patients to an exercise stress test. The primary outcome of the 6MWT, the 6-minute walk distance (6MWD), is used both to prescribe exercise training intensity (5) and as an outcome measure for CR (6).

Despite common usage of the 6MWT in CR, guidelines for the use of the 6MWT in this population were not found in a search of PubMed and CINAHL. Instead CR researchers utilize guidelines developed for pulmonary patients (6) (7) (8) (9) (10) (11) (12) (13, 14). It is unknown if the 6MWT is valid and reliable in the CR population and what changes in 6MWD would be expected following CR intervention. Repeated 6MWTs are recommended in the pulmonary (15, 16) (17) (18) and chronic heart failure populations (19) (20), although, a recent study suggests repeated 6MWTs are unnecessary in patients with heart failure who walk less than 300m on an initial 6MWT (21).

Therefore, we conducted a systematic review to examine the validity, reliability and responsiveness of the 6MWT in the outpatient CR population.

Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

Data Sources:

The first author conducted searches of OvidMEDLINE(January 1948 to March 2011), CINAHL and SPORTdiscus (January 1997 to April 2011), EMBASE (January 1980 to April 2011), Cochrane Reviews (current), and Cochrane Clinical Trials (current) databases using the search terms in Box 1. The 'date of publication' limitation varied between the databases bases upon library access availability.

Eligibility Criteria:

Trials using 6MWTs in subjects with coronary artery disease undergoing outpatient-CR were included. We included clinical trials and observational studies that described repeated 6MWTs; that compared 6MWD to established reference tests; and that examined 6MWD pre- and post-CR. We excluded trials not available as full-text articles, not in the English language, and in populations other than outpatient-CR.

Study Appraisal and synthesis methods:

The first author screened titles and abstracts of the identified articles for duplicates, and adherence to inclusion and exclusion criteria. The reference lists of the included articles were scanned for potentially relevant studies. The first author extracted and tabulated the data from the included articles under the categories of reliability, validity, and responsiveness and the third author confirmed this process.

To assess the quality of the extracted articles for reliability, validity and responsiveness, Brink and Louw's thirteen questions (22), were combined with additional criteria on responsiveness and reliability developed by Jerosch-Herold (23) and May et al (24) (25), capitalising on the strengths of each tool. Questions were adapted to ensure a yes/no response(23) (24). Studies scoring over 60% positive responses for methodology criteria relevant to the study type were considered high in quality (24) while studies scoring less than 40% were assessed as low quality. In scoring the quality of the reliability studies, we also omitted the criteria for intra- and inter-tester reliability, as the methodology design of both reliability studies, did not include these comparisons. The first and third authors, blinded to one another, undertook assessment of each article against these criteria as shown in Table 1. Differences in opinion between the two assessors were resolved through discussion and consensus.

Qualitative analysis was based upon the established level of evidence criteria in Box 2 (24) (26) and quantitative analysis was performed using Review Manager (RevMan)[a]. To enable meta-analysis of 6MWD responsiveness, the data was standardised by converting scores to mean differences and standard errors. Data from subject subgroups were included in the meta-analysis where whole of cohort figures were not available. To further explore responsiveness of the 6MWD in CR, we calculated percentage change and the effect size (27) for the subjects (and sub-groups) of the studies based on the mean and standard deviations for 6MWDs provided for the groups. We recognise that the percentage change calculated is a grouped measure and may not accurately reflect the true percentage change for the cohorts. Further, we examined the data for evidence of the ability of the 6MWD to discriminate between subjects based upon physiological factors.

Results:

The search yielded 175 acceptable articles. Figure 1 outlines the flow of article selection for analysis and the reasons for exclusions. The size of study cohorts varied, with large subject numbers in retrospective studies (6) (28) (29) (30) and smaller numbers in prospective trials (7)

(11) (31) (32). One article described a prospective study and made comparisons retrospectively with data from patient files (10).

Table 1 shows the quality appraisal of the included articles. Some studies were assessed under more than one category, i.e. reliability, validity or responsiveness, and the methodology quality of these studies varied depending upon assessment against the appraisal criteria required for each study type. One high quality study was identified for reliability, six for validity and eleven for responsiveness.

Reliability:

Qualitative analysis of the reliability data was conducted as only two studies were identified, precluding quantitative analysis. These two studies, shown in Table 2, examined the repeatability of the 6MWT in CR (7) (10) without addressing either intra-tester or inter-tester reliability (23). One study (10), assessed as high in quality, demonstrated a strong test-retest reliability (Intraclass correlation = 0.97) between repeated 6MWDs however, a statistical difference was found for repeated tests between 6MWD1 and 6MWD2 (7, 10) and between 6MWD 2 and 6MWD3 (10). These differences were 3% to 8% between the 6MWD1 and 6MWD2 (7) (10) and of 2% between 6MWD 2 and 6MWD3 (10).

Validity:

Qualitative analysis of the validity data was conducted as the variety of reference tests used for comparison with the 6MWT prevented quantitative analysis. Eight articles assessed criterion-related validity of the 6MWT; however, the predictive, diagnostic and prognostic validity of this test was not evaluated in any of the studies identified. While only two studies indicated intent to assess the validity of the 6MWT, or clinical observations during the 6MWT (10) (12), six articles used the 6MWT as part of patient assessment in CR (6) (7) (8) (9) (11) (31). Studies compared the 6MWT with a range of reference tests, the most common of which were measurements taken during a symptom-limited-exercise-test (7) (8) (9) (10) (11) (12) and the physical function subscales of quality-of-life-questionnaires (6) (10, 31). Pearson's correlation coefficient was reported in many of these studies; however the 95% confidence limits for these coefficients were not reported.

Table 3 shows the details and limitations of the study designs with the comparisons and associated statistical analyses performed. The correlation of the 6MWD to maximum power, oxygen uptake and maximum metabolic equivalents during symptom-limited-exercise-test was moderate to high ($r=.56$ to $.93$) in the four articles reporting this data. The highest correlation was reported in a study we assessed to be of moderate quality (11) examining maximum power. The second highest correlation ($r= 0.69$, $p<0.001$) was reported in a high quality study (10) which found a linear relationship between maximum metabolic equivalents achieved at a symptom-limited-exercise-test (cycle or treadmill). Both these studies used the best of repeated 6MWDs, the former the best of four (11) and the latter the best of three (10). All other studies reported on comparisons with a single 6MWD.

Two trials examined peak heart rate in symptom-limited-exercise-tests and in 6MWTs, one moderate quality study reported a moderate correlation ($r=.64$, $p <.009$) (11) and the other, assessed as high in quality (8) reported that 6MWT peak heart rate was 78%, SD=6%, of peak heart rate during symptom-limited-exercise-tests. One high (8) and one moderate (7) quality study reported a non-significant difference for heart rate assessed at ventilatory threshold and during the 6MWT. The latter study also reported a correlation of $r=.56$ between peak oxygen uptake during symptom-limited-exercise-tests and the 6MWD (7). Heart rate recovery was found

Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

to be similar following symptom-limited-exercise-tests and 6MWTs but this report failed to perform correlation analysis (12). A moderate quality study (11) reported a moderate correlation (r=.52, p <.038) between blood pressure responses during symptom-limited-exercise-tests and 6MWTs.

Three moderate to high quality studies used quality-of-life-questionnaires as reference tests for validation of the 6MWT. Two studies reported moderate correlations (r=.54 to .62, p<.001) (31) (10) while the third study found no relationships with health and function domains (6).

Responsiveness:

The ability of the 6MWT to detect change in clinical status over time and difference between subject variables was explored in 12 studies. Four studies compared different cohorts based upon sex, age, left ventricular ejection fraction and initial 6MWD(5) (28) (30) (33). One study reported an effect size for pre-CR to post-CR (6). Eleven studies reported data for pre-CR, post-CR 6MWDs and standard deviations (SD) or mean difference and SDs for inclusion in the quantitative analysis (see Table 4).

We performed meta-analysis of the data for 6MWD responsiveness using random effects analysis due to a high heterogeneity (60%) in the data. Figure 2 reports the grouped estimate of the mean difference in 6MWD pre-CR to post-CR of 60.43m, with 95% confidence intervals of 54.57m and 66.30m, p<.00001. Table 4 also shows the values we calculated for effect sizes and percentage change for the cohorts in italics. Effect sizes ranged from 0.40 to 4.28 (mean = .96; median =0.65) and the percentage change post-CR ranged from 10% to 28%. Pre-CR and post-CR 6MWD ranged from 301 to 489m and 377 to 554.5m, respectively.

Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

Discussion:

The 6MWT is a common outcome measure in CR, however, to our knowledge, we are the first to systematically review the reliability, validity and responsiveness of the 6MWT in this population. Utilizing meta-analysis, we found strong evidence that the 6MWT is responsive to change in clinical status following CR and report a grouped estimate of the mean difference in 6MWD following CR of 60.43m with 95% confidence interval of 54.57 to 66.30m. We found moderate qualitative evidence for the repeatability of the 6MWT in the CR population, with moderate to high correlation between the 6MWD and maximum metabolic equivalents achieved during symptom-limited-exercise-test and for a relationship between peak heart rate during the 6MWT and at ventilatory threshold during cycle-exercise. Further, there is moderate evidence that a learning effect occurs between repeated-6MWTs.

Reliability:

Examining reliability, we found moderate evidence for the repeatability of the 6MWT in the CR population (2 trials, 103 subjects), demonstrating a 2% to 8% change with test-retest with a .97 ICC. Repeatability was assessed with 2-repeated (7), or 3-repeated (10) 6MWTs, however, the timing of the repeated 6MWTs was not stated clearly in either study, and may have been separated several days (7, 10). We found no evidence for inter-tester or intra-tester reliability of this test in CR (zero trials, zero subjects).

Validity:

Assessing the evidence for validity, we found moderate evidence of a moderate to high correlation between the 6MWD and maximum-metabolic-equivalents achieved on symptom-limited-exercise-test (2 trials, 175 subjects) and moderate evidence for a relationship between peak heart rate responses during 6MWTs and during exercise-tests at ventilatory threshold (2 trials, 55 subjects). We found limited evidence for comparisons with all other aspects of the exercise-tests, including, peak heart rate responses during symptom-limited-exercise-tests and 6MWT (2 trials, 40 subjects), 6MWD and maximum power on symptom-limited-exercise-test (1 trial, 10 subjects); blood pressure responses in 6MWTs and those in symptom-limited-exercise-tests (1 trial, 10 subjects); 6MWD and oxygen uptake during 6MWT and during symptom-limited-exercise-tests (1 trial, 25 subjects).

We found conflicting evidence for the the relationship between 6MWD and quality-of-life-measures with moderate correlations reported in two trials (113 subjects) and no relationship found in another trial (630 subjects). The difference in results could relate to differences in subject numbers, the timing of comparisons made or to the tools used. Verrill (6) compared pre- and post-CR quality-of-life-measures and 6MWD while one study compared Short-Form-36 to pre-CR 6MWD alone (10) and the other study, does not describe the timing of the comparison (31). The studies reporting moderate correlations used the Short-Form-36-Health-Survey-physical-score (31) (10) while the other used subscales of the Ferrans-and-Powers'-Quality-of-Life-Index-Cardiac-Version-III (6). Others have reported greater body weight loss and improvement in Short-Form-36-Health-Survey-physical-score in subjects with a 23% or greater increase in 6MWD post-CR when compared to those with 12% or less improvement (28). However, while this could be a factor the two studies reporting 6MWD changes both demonstrated a 15% improvement (31) (6).

Responsiveness:

Strong evidence of the ability of the 6MWT to demonstrate change in clinical status was indicated. We report a mean difference in 6MWD following CR intervention (12 trials, 2487 subjects) of 60.43m with 95% confidence interval of 54.57 to 66.30m and a percentage or relative increase of 10% to 28% (3) (5) (6) (8) (12) (28) (29) (30) (31) (32) (33). Importantly, this relative increase was 1 to 14 times greater than the increase that would be expected by test-retest familiarization alone (7) (10). The mean difference in 6MWD reported in our analysis is higher than the minimal-clinically-important-difference of 25 to 27m reported for 6MWD in subjects with coronary artery disease (9) and subjects with chronic obstructive airways disease (33) and may reflect a good response to CR training. To our knowledge, we are the first to report 6MWDs pre-CR and post-CR across a number of studies, providing a range of expected pre-CR, post-CR 6MWDs for comparative purposes. Moreover, we have calculated effect sizes to inform relative comparisons across programs. The largest effect size of 4.28 was reported in a subgroup (28) demonstrating a narrow standard deviation compared to other groups (see Table 4 and Figure 2).

We explored the high heterogeneity of the data reported in our review and found that when data from two studies (9) (28) and Maniar's aged-less-than-65-years subgroup (30) were removed from the analysis the heterogeneity reduced to zero. These studies demonstrated a high mean difference in 6MWD pre-CR to post-CR (69.3 to 73.9m) and it is possible that differences in exercise dose between studies are the cause of this heterogeneity. Two of these studies based training prescription on symptom-limited–exercise–tests and had high exercise doses, one reporting 1.5 hours (9) training sessions and another of up to 12 months training (28), however, Gremeaux's earlier article (8) reports on 2-hour training sessions and removing this earlier article from the analysis increased heterogeneity. Another contributing factor could be the potential to improve. Maniar's aged-less-than-65-years subgroup (30) had higher BMI than all other study cohorts and poorer waist circumference, diastolic blood pressure, glucose control, and lipid profiles than Maniar's aged-greater-than-65-years subgroup. It is possible that exercise training in this subgroup had greater ability to improve 6MWD than in other groups. It is interesting however, that removing the Tallaj's subgroup with left-ventricular-ejection-fraction-less-than-40% (5), this being the group that demonstrated the greatest improvement in 6MWD, reduced the mean 6MWD difference by approximately 1m without affecting the 60% heterogeneity. This could be due to the small numbers in this subgroup. Therefore, the reasons for the heterogeneity reported remain inconclusive.

We found limited evidence for the ability of 6MWD to discriminate between subjects in CR based upon sex (1 trial, 630 subjects) and age of the subjects (1 trial, 685 subjects). Further, sex and age (29) were reported as independent predictors of change in 6MWD following CR in a multivariate linear regression analysis. These findings are consistent with reported effects of sex, weight, height and age on 6MWD in linear regression models for healthy adults (34) (35). One study (29) (156 subjects, 49 women) reported symptoms of dyspnoea and angina were more common in women than men (48% versus 24%, p = 0.007).

Limited evidence was found to support a relationship between 6MWD improvement post-CR and initial 6MWD (1 trial, 425 subjects) or left-ventricular-function (1 trial, 179 subjects). Our review excluded articles that examined heart failure subjects specifically; however, one study we included grouped CR subjects based upon degree of left-ventricular-ejection-fraction and examined the differences between these groups (5). Subjects with left-ventricular-ejection-fraction-less-than-40% increased 6MWD post-CR by 74m (26%) and had an initial 6MWD similar to another subgroup aged-75-years-or-older (286m,SD=88.7m and 283, SD=88m, respectively) (30). It could be that subjects recording low initial 6MWDs have greater room for improvement than those walking greater distances and that a “ceiling effect” may occur for 6MWD in those CR subjects who have greater initial cardiovascular fitness (28).

Limitations:

Although an extensive search of the databases was conducted it could be that by refining our search question to ensure capture of studies in the outpatient CR population and to exclude those in other populations, that we missed relevant articles. Further, in reviewing the the titles in the reference lists of retrieved full-text articles for the terms '6MWT', '6MWD' and 'outpatient CR' we may have excluded relevant articles. We did not contact authors to seek missing data or to request access to unpublished studies. In addition, our adapted quality assessment tool for reliability, validity and responsiveness may not be valid and could have led us to misrepresent the quality of the articles assessed.

A major limitation of the studies reviewed was that many examined validity using cycle-exercise-tests as the gold standard reference test. An improved capacity in walking may not result in an improved capacity in cycling due to specificity of training (36). The use of a treadmill-exercise-test as a reference test may have improved the comparisons with the 6MWT.

In addition, the studies did not consistently note the elapsed time between the reference tests and the 6MWT, the time from cardiac event to assessment, and left-ventricular-ejection-fraction for subjects. While we excluded studies in the heart failure population, we acknowledge that many subjects within the included studies will have had a degree of heart failure and the presence of heart failure may have affected subjects' response to exercise training and 6MWD.

Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

Conclusion and implications of key findings

The implication of our review is that the 6MWT has been shown to be suitable for pre- and post-CR outcome assessment, despite a learning effect of 2-8% with repeated tests. To allow program effectiveness comparisons we have reported an estimated change in 6MWD following CR of 60.4m with a .65 median effect size.

The evidence for validity against symptom-limited and ventilatory threshold exercise-tests and against quality-of-life-measurement remains inconclusive. We recommend future research examines intra-rater and inter-tester reliability of the 6MWT, further assesses the validity of the 6MWT against treadmill-exercise-tests and explores the prognostic and predictive value of 6MWDs in the CR population.

Suppliers

a. Review Manager (RevMan) [Computer program]. Version 5.1. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2011.

References:

1. Gibbons R. ACC/AHA 2002 Guideline Update for Exercise Testing: A Report of the American College of Cardiology/American Heart Association Task Force on practice Guidelines (Committee on Exercise Testing). JACC2002;40:1531-40.
2. Goble AJ, Worcester M. Best Practice Guidelines for Cardiac Rehabilitation and Secondary Prevention 1999.
3. Wright DJ, Khan KM, Gossage EM, Saltissi S. Assessment of a low-intensity cardiac rehabilitation programme using the six-minute walk test. Clin Rehabil2001 Apr;15(2):119-24.
4. (SIGN) SIGN. "Cardiac Rehabilitation – a national clinical guideline" Edinburgh Scottish Intercollegiate Guidelines Network (SIGN) & British Association for Cardiac Rehabilitation; 2002. p. 8-13.
5. Tallaj J, Sanderson B, Breland J, Adams C, Schumann C, Bittner V. Assessment of functional outcomes using the 6-minute walk test in cardiac rehabilitation: comparison of patients with and without left ventricular dysfunction. J Cardiopulm Rehabil2001 Jul-Aug;21(4):221-4.
6. Verrill D, Barton C, Beasley W, Lippard M, King C. Six-minute walk performance and quality of life comparisons in North Carolina cardiac rehabilitation programs. Heart Lung2003 Jan-Feb;32(1):41-51.
7. Gayda M, Temfemo A, Choquet D, Ahmaidi S. Cardiorespiratory requirements and reproducibility of the six-minute walk test in elderly patients with coronary artery disease. Arch Phys Med Rehabil2004 Sep;85(9):1538-43.
8. Gremeaux V, Deley G, Duclay J, Antoine D, Hannequin A, Casillas J. The 200-m fast-walk test compared with the 6-min walk test and the maximal cardiopulmonary test: A pilot study. . Am J Phys Med Rehabil2009;88:571–8.
9. Gremeaux V, Troisgros O, BenaÃ¯m S, Hannequin A, Laurent Y, Casillas J-M, et al. Determining the minimal clinically important difference for the six-minute walk test and the 200-meter fast-walk test during cardiac rehabilitation program in coronary artery disease patients after acute coronary syndrome. Archives of Physical Medicine and Rehabilitation2011;92(4):611-9.
10. Hamilton D, Haennel R. Validity and reliability of the 6-minute walk test in a cardiac rehabilitation population. J Cardiopulm Rehabil2000 May-Jun;20(3):156-64.
11. Kristjánsdóttir Á, Ragnarsdóttir M, Einarsson M, Torfason B. A comparison of the 6-minute walk test and symptom limited graded exercise test for phase II cardiac rehabilitation of older adults. Journal of Geriatric Physical Therapy2004;27(2):65-9.
12. Roberts E, Li F, Sykes K. Validity of the 6-minute walk test for assessing heart rate recovery after an exercise-based cardiac rehabilitation program. Physiotherapy2006;92(2):116-21.
13. Crapo R, Enright P, Zeballos Rwc. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med2002 Jul 1;166(1):111-7.
14. "ACSM's guidelines for exercise testing and prescription". 8th edition ed. Medicine ACoS, editor: ACSM: Baltimore, Lippincott Williams & Wilkins.; 2010.

15. Solway S, Brooks D, Lacasse Y, Thomas S. A qualitative systematic overview of the measurement properties of functional walk tests used in the cardiorespiratory domain. Chest2001 Jan;119(1):256-70.
16. Steele B. Timed walking tests of exercise capacity in chronic cardiopulmonary illness. J Cardiopulm Rehabil1996 Jan-Feb;16(1):25-33.
17. Guyatt G, Sullivan M, Thompson P, Fallen E, Pugsley S, Taylor D, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. Can Med Assoc J1985 Apr 15;132(8):919-23.
18. Sciurba F, Criner G, Lee S, Mohsenifar Z, Shade D, Slivka W, et al. Six-minute walk distance in chronic obstructive pulmonary disease: reproducibility and effect of walking course layout and length. Am J Respir Crit Care Med`2003 Jun 1;167(11):1522-7.
19. Riley M, McParland J, Stanford C, Nicholls D. Oxygen consumption during corridor walk testing in chronic cardiac failure. Eur Heart J1992 Jun;13(6):789-93.
20. Zugck C, Kruger C, Durr S, Gerber S, Haunstetter A, Hornig K, et al. Is the 6-minute walk test a reliable substitute for peak oxygen uptake in patients with dilated cardiomyopathy? Eur Heart J2000 Apr;21(7):540-9.
21. Adsett J, Mullins R, Hwang R, Hodgen A, Gibson E, Houlihan K, et al. Repeat six-minute walk tests in patients with chronic heart failure: are they clinically necessary? European Journal of cardiovascular prevention and rehabilitation 2011;0((00)):1-6.
22. Brink Y, Louw QA. Clinical instruments: reliability and validity critical appraisal. J Eval Clin Pract2011 Jun 20.
23. Jerosch-Herold C. An Evidence-Based Approach to Choosing Outcome Measures: a Checklist for the Critical Appraisal of Validity, Reliability and Responsiveness Studies. The British Journal of Occupational Therapy2005;68(8):347-53.
24. May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. Physiotherapy2010;96(3):179-90.
25. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. Journal of Physiotherapy2006;52(2):91--102.
26. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine (Phila Pa 1976)2003 Jun 15;28(12):1290-9.
27. Cohen J. Quantitative methods in Psychology :a power primer. Psychological Bulletin1992;112(1):155-9.
28. Araya-Ramirez F, Briggs KK, Bishop SR, Miller CE, Moncada-Jimenez J, Grandjean PW. Who is likely to benefit from phase II cardiac rehabilitation? Journal Of Cardiopulmonary Rehabilitation And Prevention2010;30(2):93-100.
29. Bittner V, Sanderson B, Breland J, Adams C, Schumann C. Assessing functional capacity as an outcome in cardiac rehabilitation: role of the 6-minute walk test. Clinical Exercise Physiology2000;2(1):19-26.
30. Maniar S, Sanderson B, Bittner V. Comparison of baseline characteristics and outcomes in younger and older patients completing cardiac rehabilitation. J Cardiopulm Rehabil Prev2009 Jul-Aug;29(4):220-9.

31. Ades P, Savage P, Cress M, Brochu M, Lee N, Poehlman E. Resistance training on physical performance in disabled older female cardiac patients. Med Sci Sports Exerc2003;35(8):1265-70.
32. Hung C, Daub B, Black B, Welsh R, Quinney A, Haykowsky M. Exercise training improves overall physical fitness and quality of life in older women with coronary artery disease Chest2004;126(4):1026-31.
33. Rejeski W, Foy C, Brawley L, Brubaker P, Focht B, Norris Jr, et al. Older adults in cardiac rehabilitation: a new strategy for enhancing physical function. Med Sci Sports Exerc2002 Nov;34(11):1705-13.
34. Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. Am J Respir Crit Care Med1998 Nov;158(5 Pt 1):1384-7.
35. Gibbons W, Fruchter N, Slaon S, Levy R. Reference values for a multiple repetition 6-minute walk test in healthy adults older than 20 years. J Cardiopulm Rehabil2001;21:87-93.
36. Astrand P, Rodahl K. Textbook of work physiology: physiological bases of exercise: international edition. Third edition. Sydney, Australia: McGraw-hill Book Co. ; 1986.

Box 1: Search question
(six min* walk or six-min* walk or 6 min* walk or 6-min* walk or 6MW* or 6-MW*) and (cardiac rehabilitation or coronary artery disease or coronary heart disease or heart valve) and (valid* or reliab* or distance or value or predict*) not (heart Failure) not (lupus) not (multiple sclerosis) not (chronic obstructive pulmonary disease or COPD or respiratory disease).

Box 2: Levels of evidence

Strong	Consistent findings from three or more high-quality studies
Moderate	Consistent findings from at least one high-quality study and a number of low-quality studies
Limited studies	Consistent findings in one or more low-quality
Conflicting	Inconsistent findings irrespective of study quality
No evidence	No studies found

Figure 1: Flow of studies in systematic review

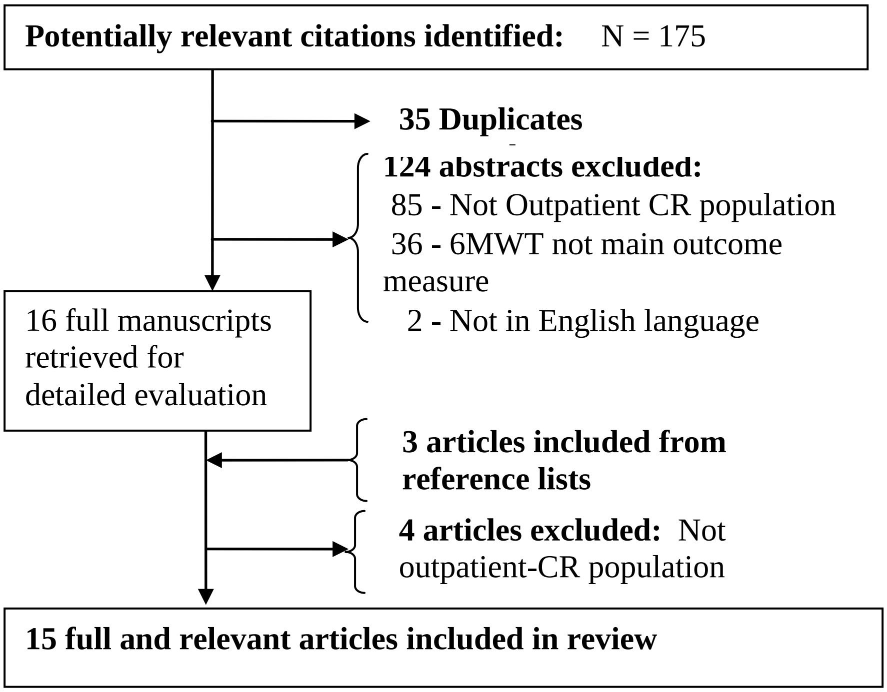
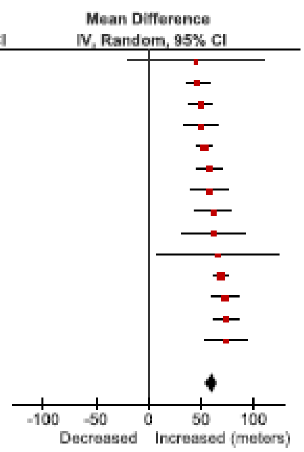


Figure 2: Forest plot of the mean difference in reported 6MWD pre-CR to post-CR

Study or Subgroup	Mean Difference	SE	Post-CR (N) Total	Pre-CR (N) Total	Weight	Mean Difference IV, Random, 95% CI
Ades 2003	45.14	32.63	19	19	0.8%	45.14 [-18.81, 109.09]
Bittner 2000	46.99	5.41	156	156	9.8%	46.99 [36.39, 57.59]
Tallaj 2001 ≥40% LVEF	49.84	5.57	129	129	9.6%	49.84 [38.92, 60.76]
Verrill 2003 Women	50.53	7.8	159	159	7.3%	50.53 [35.24, 65.82]
Maniar 2009 ≥ 65y	53.49	3.34	297	297	12.2%	53.49 [46.94, 60.04]
Verrill 2003 Men	58.19	6.22	429	429	8.9%	58.19 [46.00, 70.38]
Rejeski 2002	58.33	9.58	129	147	5.8%	58.33 [39.55, 77.11]
Wright 2001	62	8.69	159	159	6.5%	62.00 [44.97, 79.03]
Gremeaux 2009	62.7	15.09	30	30	3.1%	62.70 [33.12, 92.28]
Roberts 2006	66.2	29.15	30	30	1.0%	66.20 [9.07, 123.33]
Maniar 2009 < 65y	69.35	3.78	388	388	11.7%	69.35 [61.94, 76.76]
Gremeaux 2011	73.2	6.28	81	81	8.8%	73.20 [60.89, 85.51]
Araya-Ramirez 2010	73.9	6.34	425	425	8.8%	73.90 [61.47, 86.33]
Tallaj 2001 < 40% LVEF	74.18	9.88	50	50	5.6%	74.18 [54.82, 93.54]
Total (95% CI)			**2481**	**2499**	**100.0%**	**60.43 [54.57, 66.30]**

Heterogeneity: $Tau^2 = 62.02$; $Chi^2 = 32.23$, df = 13 ($P = 0.002$); $I^2 = 60\%$

Test for overall effect: $Z = 20.19$ ($P < 0.00001$)



Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

Table 1: Critical appraisal of articles for Reliability, Validity and Responsiveness

Questions	1	2	3[a]	4[b]	5[b]	6[b]	7[a]	8[b]	9[a]	10	11[a]	12	13	14[c]	15	16	17	% Yes
Reliability Studies scored out of 10																		
Gayda 2004	Y	N	n/a	N/A	N/A	N	n/a	N	n/a	Y	n/a	N	Y	n/a	N	N	Y	40
Hamilton 2000	Y	N	n/a	N/A	N/A	N	n/a	N	n/a	Y	n/a	Y	Y	n/a	Y	Y	Y	70
Validity Studies scored out of 12																		
Ades 2003	Y	N	N	n/a	n/a	n/a	Y	n/a	Y	N	N	Y	Y	n/a	Y	N	Y	58
Gayda 2004	Y	N	Y	n/a	n/a	n/a	N	n/a	Y	Y	Y	N	Y	n/a	N	N	Y	58
Gremeaux 2011	Y	Y	Y	n/a	n/a	n/a	Y	n/a	Y	Y	Y	Y	Y	n/a	Y	Y	Y	100
Gremeaux 2009	Y	N	Y	n/a	n/a	n/a	Y	n/a	Y	Y	Y	n/a	N	n/a	Y	N	Y	67
Hamilton 2000	Y	N	Y	n/a	n/a	n/a	N	n/a	Y	Y	Y	Y	Y	n/a	Y	Y	Y	83
Kristjánsdóttir 2000	N	N	Y	n/a	n/a	n/a	Y	n/a	Y	Y	Y	n/a	N	n/a	Y	N	Y	58
Roberts 2006	Y	Y	Y	n/a	n/a	n/a	Y	n/a	Y	Y	Y	Y	Y	n/a	Y	N	Y	92
Verrill 2003	Y	N	Y	n/a	n/a	n/a	Y	n/a	Y	Y	Y	N	Y	n/a	Y	Y	N	75
Responsiveness Studies scored out of 9																		
Ades 2003	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	N	n/a	Y	Y	Y	Y	N	Y	67
Araya-Ramirez 2010	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	N	Y	Y	Y	Y	N	67
Bittner 2000	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	N	Y	Y	Y	Y	N	67
Gremeaux 2011	Y	Y	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	Y	Y	Y	Y	Y	Y	100
Gremeaux 2009	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	n/a	N	Y	Y	N	Y	56
Hung 2004	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	N	n/a	Y	Y	Y	Y	N	Y	67
Maniar 2009	Y	Y	n/a	n/a	n/a	n/a	n/a	n/a	n/a	N	n/a	N	Y	Y	Y	Y	Y	78
Rejeski 2002	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	Y	Y	Y	Y	Y	Y	89
Roberts 2006	Y	Y	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	Y	Y	Y	Y	N	Y	89
Tallaj 2001	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	N	Y	Y	Y	Y	N	67
Verrill 2003	Y	N	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	N	Y	Y	Y	Y	N	67
Wright 2001	Y	Y	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Y	n/a	Y	Y	Y	Y	Y	N	89

No, N; Yes, Y, not applicable, n/a; grey shaded n/a, criteria not applicable for the study type; grey shaded N/A, inter- and intra-tester criteria not applicable to the study design

[a] Items relevant to validity; [b] Items relevant to reliability; [c] items relevant to responsiveness; Items unmarked were relevant to all study types.

Questions (adapted from Brink and Louw (22) Jerosch-Herold (23) and May et al (24, 25): 1, If human subjects were used, did the authors give a detailed description of the sample of subjects used to perform the (index) test?; 2, Did the authors clarify the qualification, or competence of the rater(s) who performed the(index) test?; 3, Was the reference standard explained?; 4, If interrater reliability was tested, were

Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

raters blinded to the findings of other raters?; 5, If intrarater reliability was tested, were raters blinded to their own prior findings of the test under evaluation?; 6, Was the order of examination varied?; 7, If human subjects were used, was the time period between the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?; 8, Was the stability (or theoretical stability) of the variable being measured taken into account when determining the suitability of the time interval between repeated measures?; 9, Was the reference standard independent of the index test?; 10, Was the execution of the (index) test described in sufficient detail to permit replication of the test?; 11, Was the execution of the reference standard described in sufficient detail to permit its replication?; 12, Were withdrawals from the study explained?; 13, Were the statistical methods appropriate for the purpose of the study? 14, Does the instrument capture clinical change?; 15, Were subjects selected randomly or consecutively?; 16, Were the number of subjects >50 or was a sample size calculation provided?; 17, Did subjects give consent prior to testing?

Table 2: Reliability of six-minute walk test prior to participation in cardiac rehabilitation

Reference	Sample Size	Protocol	Test 1 (SD)	Test 2 (SD)	Test 3 (SD)	% change Test 1-Test 2	% change Test 2-Test3	% change Test 1-Test 3	ICC	Study Limitation
Gayda 2004	9	Repeated 6MWTs =2, during CR.	462(78)	498(58)[a]	n/a	8	n/a	n/a	n/a	Unknown timing of 2nd 6MWT
Hamilton 2000	94	Repeated 6MWT =3, pre-CR on non-consecutive days.	521(112)[b]	539(114)[b]	550(122)	3	2	6	.97[c]	n/a

CR Cardiac Rehabilitation
ICC Intraclass correlation
NS non-significant
n/a not applicableSD Standard Deviation
[a] P<.002, Test 1 v Test 2
[b] P < 0.001, Test 1 v Test 2, Test 2 vs. Test 3
[c] ICC for Test 1 to Test 2, Test 2 to test 3

Table 3: Validity of six-minute walk test in cardiac rehabilitation population

Reference	Study Design	Time of Testing	Sex (n)	Age (SD) Years	Diagnosis	6MWT protocol	Standard Test for Comparison	Comparison / Correlation	p-value
Ades 2003	Random controlled trial: 6 months training in CR, 3 times/wk.	Pre & post CR	Men(0) Women(33) n=19 in intervention group	73 (6)	CAD	Cress 1996	SF-36 questions on 'walk a mile 'walk a block'	6MWD, r=0.54 6MWD, r= 0.38	.0001 .009
Gayda 2004	Validity Study: Expired gas analysis during SLET & 6MWT	During CR	Men (22) Women (3)	60 (10)	CAD	Guyatt 1985	6MWT peak oxygen uptake SLET peak oxygen uptake SLET Peak oxygen uptake and heart rate at VT	6MWD, r= .58 6MWD, r= .56 6MWT Peak oxygen uptake and heart rate	n/a n/a NS
Gremeaux 2011	Prospective study: using distribution and anchoring-based methods. Training 8-weeks, 3 times/wk	Pre & post CR	Men (77) Women(4)	58 (8)	CAD	ATS 2002	SLET Maximal METs on treadmill	6MWD, r= .59	<.05
Gremeaux 2009	Random cross over design: SLET followed by 200m fast-walk Test or 6MWT. 6-weeks training , 3 times/wk. 10 subjects performed SLET with gas exchange to determine VT	Pre & post CR	Men (28) Women (2)	52 (8)	CAD	ATS 2002	SLET peak heart rate, before and after CR. SLET heart rate at VT	6MWT peak heart rate, = 78% (SD = 6%) of SLET peak heart rate before and after CR 6MWT peak heart rate	n/a NS
Hamilton 2000	Validity and Reliability study: in CR phase II and III.	Pre & post CR	Men (61) Women (33)	63 (10)	CAD, NYHA Class: 1=78% 2=19% 3=2%	Steele 1996 (ACSM 1995)	DASI Physical Function subscale of SF-36 SLET Maximum METs on treadmill or cycle	6MWD, r= .502 6MWD, r= .624 6MWD, r= .687	< .001 < .001 < .001
Kristjánsdóttir 2000	Validity Study: Comparison of SLET to best of four 6MWTs at 3-months post- surgery	During CR	Men (6) Women (4)	76 (6)	CAD Post Cardiac surgery	Guyatt 1984 & 1985	SLET maximum power SLET maximum heart rate SLET maximum systolic blood pressure	6MWD, r=.93 6MWT maximum heart rate, r=.64 6MWT maximum systolic blood pressure, r=.52	< .001 < .009 < .038

Roberts 2006	Longitudinal cohort design. 6MWT and SLET pre and post 8-week CR	Pre & post CR	Men (20) Women (10)	62(9)	CAD, LVEF >30%, NYHA class 1- 3	ATS 2002	SLET Heart rate recovery significantly improved in each 30-second interval of a 2-minute recovery period	6MWT Heart rate recovery significantly improved in each 30-second interval of a 2-minute recovery period	n/a
Verrill 2003	Non-experimental, prospective, comparative in multiple CR programs across Nth Carolina	Pre & post CR	Men (424) Women(206)	61(10)	CAD 4% CCF 1% Valve Disease	Hamilton 2000	Subscales of the Ferrans and QOLI, including the Health and Function domain	6MWD = No relationships were found with any domain	n/a

American College of Sports Medicine, ACSM; American Thoracic Society, ATS; Cardiac Rehabilitation, CR; Classification, Class; Congestive Cardiac Failure, CCF; Coronary artery disease, CAD; Duke Activity Status Index ,DASI; Left Ventricular Ejection Fraction, LVEF; Metabolic equivalents, METs; Not significant, NS; New York Heart Association, NYHA; Powers' Quality of Life Index-Cardiac Version III QOLI; Six-minute walk test, 6MWT; Six-minute walk test distance, 6MWD; Short Form 36 Health Survey, SF-36; Symptom limited Exercise Test, SLET; Ventilatory Threshold,VT; Times per week, x/wk; Pearson's Correlation coefficient, r.

Bellet RN, et al. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy (2012), doi:10.1016/j.physio.2011.11.003

Table 4: Reported responsiveness of the 6MWD Pre-CR and Post-CR

Study	Subgroups	Total (N)	Men	Women	Age	SD	Pre-CR 6MW	SD	Post-CR	SD	Change	SD	% increase	SD	Effect Size [a]	Test	P-value
Ades 2003	Resistance group:	19	0	19	73	6	309	101	355	100	43	51	15	n/a	*0.45*	ANOVA	<.005
Araya-Ramirez 2010	Entire cohort:	425	293	132	62	12	399	87	472	97	74	n/a	20	16.3	*0.84*	Paired t-test	<.001
	1st Tertile:	142	n/a	n/a	n/a	n/a	301	64	383	89	n/a	n/a	28	20	*1.28*	Paired t-test	n/a/
	2nd Tertile:	141	n/a	n/a	n/a	n/a	410	18	487	57	n/a	n/a	19	14	*4.28*	Paired t-test	n/a
	3rd Tertile:	142	n/a	n/a	n/a	n/a	485	37	548	58	n/a	n/a	13	10	*1.7*	Paired t-test	n/a
Bittner 2000		156	107	49	60	11	321	102	n/a	n/a	47	68	*15*	n/a	*0.53*	Paired t-test	<.0001
Gremeaux 2011		81	77	4	58	9	n/a	n/a	n/a	n/a	73	57	16	12.2	*n/a*	n/a	n/a
Gremeaux 2009		30	28	2	52	9	490	33	552	76	63	56	*13*	n/a	*1.9*	Not stated	<.01
Hung 2004	Aerobic training :	9	0	9	70	6	418	74	n/a	n/a	n/a	n/a	10	n/a	*0.56*	ANOVA	<.05
	Aerobic & Strength training:	9	0	9	71	7	438	110	n/a	n/a	n/a	n/a	10	n/a	*0.40*	ANOVA	<.05
Maniar 2009	<65 years: [b]	388	286	102	54	7	360	97	n/a	n/a	69	74	*19*	n/a	*0.71*	Not stated	<.01
	≥65 years:	297	201	96	71	5	317	98	n/a	n/a	53	58	*17*	n/a	*0.54*	Not stated	<.01
	65-74 years: [d]	226	156	70	69	3	328	99	n/a	n/a	55	60	*17*	n/a	*0.56*	Not stated	n/a
	75 + years:	71	45	26	79	3	283	88	n/a	n/a	48	50	*17*	n/a	*0.55*	Not stated	n/a
Rejeski 2002		129	75	72	65	8	456	76	514	82	57	48	*13*	n/a	*0.76*	t-test	< 0.001
	Women:	72	0	72	n/a	n/a	426	67	486	75	60	39	*14*	n/a	*0.91*	t-test	< 0.001
	Men:	75	75	0	n/a	n/a	484	75	539	81	54	55	*11*	n/a	*0.73*	t-test	< 0.001
Roberts 2006		30	20	10	62	9	486	114	553	112	n/a	n/a	*14*	n/a	*0.59*	Paired t-test	<.001
Tallaj 2001	LVEF < 40%: [e f]	50	33	17	61	10	286	89	n/a	n/a	74	70	*26*	n/a	*0.84*	Paired t-test	<.01
	LVEF ≥40%:	129	84	45	61	11	329	99	n/a	n/a	50	63	*15*	n/a	*0.5*	Paired t-test	<.01
Verrill 2003	Women: [g]	206	0	206	n/a	n/a	328	80	379	79	n/a	n/a	*15*	n/a	0.63	Not stated	<.001
	Men:	424	424	0	n/a	n/a	386	90	444	92	n/a	n/a	*15*	n/a	*0.65*	Not stated	<.001
Wright 2001		159	103	56	62	9	315	76	377	79	62	53	*20*	n/a	*0.82*	Paired t-test	< .001

Cardiac Rehabilitation, CR; Meters, m; Standard Deviation, SD, Not Significant, NS, Not applicable, n/a; six-minute walk test distance, 6MWD. Age is in mean years; Pre-CR, Post-CR and Change in 6MWDs are mean distances in meters (6MWDs, where applicable were converted from feet to meters); P-Values are for Pre-CR to Post-CR comparisons of 6MWD. *Italics* indicate data calculated for this review.

a Effect Size was calculated by the authors using published data within the study and Cohen's formula, except Verrill (2003) where effect size was reported
b p<.001 between groups for 6MWD in subjects aged < 65 years versus subjects aged ≥ 65 years at pre-CR assessment, t-test
c p<.01 between groups for 6MWD in subjects aged < 65 years versus subjects aged ≥ 65 years at post-CR assessment, t-test
d p<.001 between groups for 6MWD in subjects aged 65-74 years versus subjects aged 75+ years at pre-CR assessment, t-test
e p<.008 between groups for 6MWD in subjects with LVEF < 40% versus subjects with LVEF ≥40% at pre-CR assessment, unpaired t-test
f p<.07 between groups for 6MWD in subjects with LVEF < 40% versus subjects with LVEF ≥40% at post-CR assessment, unpaired t-test
g p<.0001 between groups for 6MWD in men versus women, at pre-CR assessment and at post-CR assessment, ANOVA for repeated measures