Reliability of one-repetition maximum performance in people with chronic heart failure.

Abstract

Purpose: To evaluate the intra-rater and inter-rater reliability of the one-repetition maximum strength test in people with chronic heart failure.

Design: Intra-rater and inter-rater (within-therapist and between-therapist) reliability study.

Setting: A public tertiary hospital in northern metropolitan Melbourne.

Participants: 24 participants (nine female, mean age 71.8 ±13.1 years) with mild to moderate heart failure of any aetiology.

Methods: Lower limb muscle strength was assessed by determining the maximum weight that could be lifted once through range and with proper technique using a leg press. Intra-rater reliability was tested by one assessor completing this measure on two separate occasions (two - five days apart). Inter-rater reliability was tested by two assessors in random order.

Statistical Analyses: Intra-class correlation coefficients and associated 95% confidence intervals were calculated. Bland and Altman analyses were also conducted, including calculation of mean differences between measures ($\bar{d}$) and limits of agreement (1.96 x Standard Deviation $_{\text{difference}}$).

Results: 10 intra-rater and 21 inter-rater assessments were completed. Excellent intra-rater (Intra-class correlation coefficient $_{2,1}$ 0.96) and inter-rater (Intra-class correlation coefficient $_{2,1}$ 0.93) reliability was found. Intra-rater assessment showed less variability (mean difference 4.5 kg, limits of agreement -8.11 to 17.11 kg) than inter-rater agreement (mean difference -3.81 kg, limits of agreement -23.39 to 15.77

kg).

Conclusion: One-repetition maximum determined using a leg press is a reliable measure in people with heart failure. Given the larger limits of agreement for inter-rater reliability, assessment by the same rater on each testing occasion is recommended.

Keywords

Chronic heart failure, strength, 1-RM, one-repetition maximum, reliability

Introduction

Chronic heart failure affects 26 million people worldwide [1] at an estimated annual cost of over $1 billion in Australia [2]. People with heart failure experience exertional dyspnoea, fatigue and weakness [3] leading to reduced exercise tolerance. The mechanisms behind limitations in physical activity among people with heart failure include inadequate blood flow to skeletal muscles, inability to increase cardiac output in response to physical activity, [3] muscle weakness [4] and muscle atrophy [5]. Reduction in muscle function contributes significantly to exercise intolerance in people with chronic heart failure and its cause is multifactorial [6]. One study reported quadriceps strength was the most important individual correlate of exercise tolerance in people with chronic heart failure, [4] while another study found quadriceps weakness was predominately due to loss of muscle mass and suggested exercise tolerance was significantly affected by muscle atrophy [5].

To minimise the effect of muscle atrophy and increase muscle strength, exercise is a recommended component of heart failure rehabilitation [3]. To ascertain if treatment is successful, therapists require reliable outcome measures that are easily used in the clinic. To date, limited studies [7-9] have investigated the reliability of muscle strength outcomes in patients with chronic heart failure with complex dynamometry equipment primarily being used. This is consistent with the literature available for other chronic disease such as chronic obstructive pulmonary disease [10] and chronic stroke [11], where only the reliability of dynamometry has been investigated, with the exception of one study that explored the reliability of an estimated one-repetition maximum (1-RM) in people with Type 2 diabetes [12]. Although isokinetic dynamometry is reliable and considered as the gold standard, with one study suggesting that compared to isokinetic dynamometry the use of the 1-RM technique overestimates strength gains over time, [7] this type of equipment is expensive and not commonly available in regular clinics.

In healthy adults 1-RM testing has also demonstrated good reliability [13] and completing a 1-RM measurement requires common gymnasium equipment, and so with no previous studies having assessed the reliability of the 1-RM strength measure in people with heart failure or in fact chronic disease, this study aimed to determine both the intra-rater and inter-rater reliability of 1-RM with a leg press in people with mild to moderate chronic heart failure.

We hypothesised the leg press 1-RM in people with chronic heart failure would demonstrate good intra-rater and inter-rater reliability.

Methods

Research design: This was an intra-rater and inter-rater (within-therapist and between-therapist) reliability study. For inter-rater reliability the order of testing by the two assessors was randomly generated using a random list generator [14]. Ethics approval for the study was obtained from the relevant hospital and university human ethics committees.

Participants: This study was conducted alongside a randomised controlled trial investigating the effects of eccentric exercise in people with chronic heart failure (ClinicalTrials.gov Identifier: NCT02223624). The eligibility criteria for this reliability study were the same as for the randomised controlled trial. Patients were included if they were: (1) aged 18 years or above; (2) had a clinical diagnosis of mild to moderate heart failure (any aetiology); (3) were medically stable; and (4) had been assessed by a physiotherapist as having no contraindications to exercise. Where there were concerns about an individual taking part, clearance was sought from the treating cardiologist.

The exclusion criteria were: (1) hospitalisation for an exacerbation of chronic heart failure within the previous month; (2) severe heart failure classified as level four on the New York Heart Association classification (i.e. short of breath at rest); (3) a concurrent unstable medical condition such as uncontrolled angina, diabetes or hypertension; (4) dementia or a psychological disorder that would interfere with participation in group exercise; (5) participation in a cardiac or heart failure rehabilitation program in the previous six months; (6) the presence of a

contraindication to exercise or (7) the presence of any pre-existing neurological or musculoskeletal condition, for example stroke, that on assessment was deemed to interfere with exercise participation.

Participants were recruited following referral to heart failure rehabilitation either from local general practitioners, a heart failure clinic or referral from an acute hospital admission at a metropolitan health service located in the north of Melbourne.

1-RM leg press testing protocol: All assessments were completed in an air-conditioned gymnasium of a hospital. Assessments were conducted at the same time of day (between 10:00 a.m. and 1:00 pm). Intra-rater reliability was measured by one assessor completing testing on two separate occasions two - five days apart. Inter-rater reliability was completed by two assessors in random order with a short rest period (five -10 minutes) in between. Both inter-rater assessors were physiotherapists with multiple years of experience. Assessor one had 11 years of clinical experience, including experience completing heart failure assessments and rehabilitation. Assessor two had five years of clinical experience including experience working with people with heart failure during acute hospitalisation. Neither assessor had previously used the 1-RM assessment of leg strength as an outcome measure for heart failure rehabilitation. All intra-rater assessments were completed by the same physiotherapist (assessor two).

A multi-gym leg press apparatus (ACUFIT ENTERPRISE Co., LTD, Taiwan) was used to perform the testing. Participants were instructed on correct performance by the assessor. This involved sitting upright on the leg press apparatus with their back against the support. Feet were placed flat on the platform, shoulder width apart and with neutral rotation. The seat was moved forward or back to create 90 degrees knee flexion which was measured using a goniometer. Participants were prompted to place their hands on the hand grips at their side. Participants were instructed to straighten their knees, moving slowly through range until extended fully (but not hyperextended).

Once correct posture was obtained (using demonstration if necessary) participants warmed up by completing five -10 submaximal (~50% maximum) repetitions. Following this, the assessor estimated an initial near maximum load. Rest periods of three - five minutes between attempts were allowed. Assessors progressed resistance by 5 kg each attempt (1 plate) or 5% whichever was greater and aimed to determine 1-RM within four attempts. The final weight successfully lifted through full range of motion was recorded as the 1-RM. This procedure was based on that described by the American College of Sports Medicine and National Strength and Conditioning Association [15, 16]. Participants were not informed of their results throughout the procedure.

Statistical Analyses: Based on a calculation by Walter, Eliasziw and Donner, [17] in order to achieve an Intra-Class Correlation (ICC) value of greater than 0.8, a sample size of n = 46 was required, assuming an alpha of 0.05 and power of 80%. Although some suggest a level of agreement of 0.7 is good [18] based on previous

reliability studies in the heart failure population the minimum value of 0.8 was deemed clinically acceptable [8]. During the completion of the study, assessor two moved overseas and so the decision was made to cease further assessments rather than recruiting a new assessor to avoid introducing a new source of variability. This meant that the estimated sample size was not reached.

Intra-class correlation coefficients ($ICC_{2, 1}$) and the associated 95% confidence interval (CI) were calculated. Bland and Altman plots were used to assess agreement between testing occasions, which involved calculation of the mean difference between measures ($\bar{d}$) and the limits of agreement (1.96 x SD_{diff}).

Results

Participants: The sample consisted of 24 participants with chronic heart failure (nine female) and mean age 72 ± 13 years (table 1). Heart failure severity based on New York Heart Association classification was mild to moderate (n = 13 class II). Sixteen participants had systolic dysfunction on echocardiogram. Four participants had diastolic dysfunction, two had evidence of both, one had no reported dysfunction on echocardiogram and the final participant was newly diagnosed and awaiting echocardiogram, thus diagnosis was based on clinical presentation. Ejection fraction (n=19) was reduced with a mean percentage of 37.0 ± 13.5. Four participants did not have a documented ejection fraction on transthoracic echocardiogram and one participant was awaiting echocardiogram. Cause of heart failure was classified as ischaemic in 10 participants and non-ischaemic (including

valvular) in 13 participants. All participants were taking cardiac medications (table 1).

Exercise protocol: Fourteen participants could not achieve the starting position of 90 degrees of knee flexion, due to reduced range of motion or body stature. The actual starting position ranged from 45 to 90 degrees (mean assessor one 85±11 degrees, mean assessor two 82 ± 9 degrees).

Two participants were able to lift the maximum possible weight (120 kg). All participants could lift the minimum weight of 5 kg.

Reliability: The inter-rater ICC $_{(2,1)}$ was 0.93 (95% CI 0.83 – 0.97) (table 2) suggesting an excellent level of agreement. The intra-rater ICC $_{(2,1)}$ was also excellent at 0.96 (95% CI 0.81 – 1.00). The Bland and Altman method showed a mean difference between measures ($\bar{d}$) for inter-rater reliability (figure 1) of -3.81 kg with limits of agreement of -23.39 to 15.77 kg. For intra-rater reliability (figure 2) the mean difference between measures ($\bar{d}$) was 4.5 kg and limits of agreement were -8.11 to 17.11 kg.

Adverse events: One participant reported chest discomfort during testing that quickly resolved with rest. No other negative events were reported.

Discussion

Our results suggest assessment of lower limb strength using leg press 1-RM had excellent reliability in people with heart failure. The mean difference between

testing occasions for both inter-rater reliability and intra-rater reliability was small, equivalent to less than one 5 kg plate on the leg press. However, the limits of agreement were wide, particularly for inter-rater assessments (-23.39 kg to 15.77 kg), with the range of inter-rater differences varying from no difference to 25 kg difference between testing occasions. As a result, the use of one outcome assessor is recommended to accurately measure strength changes using this method in people with heart failure. Increasing evidence suggests that not only is lack of muscle strength a significant result of chronic heart failure [4, 5] but that strengthening exercises can safely be included in rehabilitation programs to address this [3]. The results of this study allow staff and patients with chronic heart failure completing strengthening exercise to assess baseline strength, which can then be used in exercise prescription and tracking of progress. Given the lack of studies demonstrating the reliability of this technique in other chronic diseases but its frequent use as an outcome to measure strength changes, this present study may have applications across a broader population. Further research in other chronic disease populations is recommended.

In previous reliability studies including people with heart failure and other cardiac conditions, familiarisation sessions were reported to improve reliability [9]. In our study, for five of the 21 inter-rater assessments the participant had completed the procedure once before. The mean difference for these five assessments was -12 kg compared with the mean for all measurements of -3.81 kg suggesting that, familiarisation did not help in increasing agreement. When intra-rater reliability was investigated using the Bland and Altman plot it revealed that in nine out of 10 assessments the participant scored higher values than on the first. This suggests a

learning effect, either on the part of the patient or on the part of the assessor, who may feel more confident when completing the procedure with a participant for the second time in what has been generally considered a high-risk population. No consistent differences were observed between assessor one and assessor two, suggesting that level of experience did not affect results.

Limitations

The apparatus used in this study measured strength in 5 kg increments which limits sensitivity to change and affects the analysis of reliability. Contrary to expectations, two participants experienced a ceiling effect by being able to lift the maximum possible weight (120 kg). While one participant was young (49 years) with mild, poorly defined heart failure (New York Heart Association class 1) the other participant was assessed as New York Heart Association class 3 and had dilated cardiomyopathy with severe systolic dysfunction and an ejection fraction of 16%. This supports the finding that heart function particularly as measured by ejection fraction does not necessarily correlate with exercise tolerance [19] and perhaps participant age was a greater determinate of strength (40 year old male). One participant was only able to lift the minimum amount, however, a floor effect was not evident. The assessment procedure required both assessors aim for a set-up position of 90 degrees of knee flexion. This was not achievable for 14 (58%) of participants suggesting it may need to be revised.

The major limitation identified for this study is that the previously determined sample size was not reached due to the departure of assessor two from the country.

Introducing a new assessor would have increased variability, rather than increasing the confidence in the results and as such was not carried out. Given the high $ICC_{(2,1)}$ achieved and the spread of values spans across the range of possible values on the Bland and Altman plot it is likely that the sample size was sufficient to demonstrate reliability however a further study with larger sample should be considered.

Conclusion

One-repetition maximum determined using a leg press is a reliable measure in people with heart failure. Given the larger limits of agreement for inter-rater reliability, assessment by the same rater on each testing occasion is recommended.

Declaration of interest

The authors report no conflicts of interest.

References

1.   Ambrosy AP, Fonarow GC, Butler J, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol. 2014;63(12):1123-33.

2.   National Heart Foundation of Australia. Heart Failure Factsheet. [Internet] Heart Foundation [Cited 2016 Nov 30] Available from:
https://www.heartfoundation.org.au/images/uploads/publications/HF-Fact-sheet-02-Heart-Failure.pdf

3.   National Heart Foundation of Australia and the Cardiac Society of Australia and New Zealand -Chronic Heart Failure Guidelines Expert Writing Panel. Guidelines for the prevention, detection and management of chronic heart failure in Australia.[Internet] Updated October 2011. [Cited 2013 Jul 1] Available from:
http://www.heartfoundation.org.au/sitecollectiondocuments/chronic_heart_failure_guidelines_2011.pdf.

4.   Volterrani M, Clark AL, Ludman PF, et al. Predictors of exercise capacity in chronic heart failure. Eur Heart J. 1994;15(6):801-9.

5.   Harrington D, Anker SD, Chua TP, et al. Skeletal muscle function and its relation to exercise tolerance in chronic heart failure. J Am Coll Cardiol. 1997;30(7):1758-64.

6.   Gosker HR, Wouters EF, van der Vusse GJ, et al. Skeletal muscle dysfunction in chronic obstructive pulmonary disease and chronic heart failure: underlying mechanisms and therapy perspectives. Am J Clin Nutr. 2000;71(5):1033-47.

7.   Feiereisen P, Vaillant M, Eischen D, et al. Isokinetic versus one-repetition maximum strength assessment in chronic heart failure. Med Sci Sports Exerc. 2010;42(12):2156-63.

8.   Quittan M, Wiesinger GF, Crevenna R, et al. Isokinetic strength testing in patients with chronic heart failure--a reliability study. Int J Sports Med. 2001;22(1):40-4.

9.   Selig SE, Carey MF, Menzies DG, et al. Reliability of isokinetic strength and aerobic power testing for patients with chronic heart failure. J Cardiopulm Rehabil. 2002;22(4):282-9.

10.   O'Shea SD, Taylor NF, Paratz JD. Measuring muscle strength for people with chronic obstructive pulmonary disease: retest reliability of hand-held dynamometry. Arch Phys Med Rehabil. 2007;88(1):32-6.

11.   Ekstrand E, Lexell J, Brogardh C. Isometric and isokinetic muscle strength in the upper extremity can be reliably measured in persons with chronic stroke. J Rehabil Med. 2015;47(8):706-13.

12.   Abdul-Hameed U, Rangra P, Shareef MY, Hussain ME. Reliability of 1-Repetition Maximum Estimation for Upper and Lower Body Muscular Strength Measurement in Untrained Middle Aged Type 2 Diabetic Patients. Asian J Sports Med. 2012;3(4):267-73.

13.   Verdijk LB, van Loon L, Meijer K, et al. One-repetition maximum strength test represents a valid means to assess leg strength in vivo in humans. J Sports Sci. 2009;27(1):59-68.

14.   Dallal GE. Welcome to Ransimization. [Internet] 2013 Dallal G [Cited 2014 Jun 1] Available from: http://www.randomization.com/

15.   Baechle TR, Earle RW. National Strength and Conditioning Association-Essentials of Strength Training and Conditioning. Third edition: Human Kinetics; 2008. p. 398.

16.   Thompson WR, Gordon NF, Pescatello LS. ACSM's Guidelines for Exercise Testing and Prescription. Eighth edition. US: Wolters Kluwer Health. Lippincott Williams & Wilkins; 2010. p. 90.

17.   Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17(1):101-10.

18.   Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Inj. 2011;42(3):236-40.

19.   Franciosa JA, Park M, Levine TB. Lack of correlation between exercise capacity and indexes of resting left ventricular performance in heart failure. Am J Cardiol. 1981;47(1):33-9.

20.   World Health Organization. BMI classification [Internet] 2006 World Health Organization. [Cited 2016 Nov 11] Available from: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html

Figure 1: Bland and Altman plot showing mean measurements against differences for inter-rater reliability, including levels of agreement (LOA).

Figure 2: Bland and Altman plot showing mean measurements against differences for intra-rater reliability, including levels of agreement (LOA).

Table 1: Demographic Data for 1-RM participants

Characteristic	Inter-rater reliability (n=21)	Intra-rater reliability (n=10)
Sex (male/female)	13/8	8/2
Mean age (SD) (y)	71.9 (13.9)	71.3 (11.8)
Language (English/non-English speaking) (number)	15/6	7/3
Mean height (SD) (cm)	158.2 (39.1)	168.0 (11.6)
Mean weight (SD) (kg)	87.8 (23.2)	97.8 (21.8)
Mean BMI (SD) (kg/m2)*	32.1 (6.2)**	34.6 (7.0)**
NYHA Classification (Class 1-3) (number)	8/10/3	3/7/0
Diagnosis (Systolic/ Diastolic/ combined heart failure) (number)	14/4/2 (1 unreported)	8/1/0 (1 unreported)
Aetiology (Ischaemic/ Non-ischaemic heart failure) (number)	8/12 (1 awaiting investigation)	4/6
Mean EF (SD) (%)	36.0 (13.9)***	38.3 (13.9)***
Medications (number) Beta blocker	17	7
ACE inhibitor	14	7
Calcium channel blocker	5	3
Nitrate	4	2
Diuretic	15	5
Statin	14	8
Anticoagulant	18	10
ARA 2	5	2
Aldosterone antagonist	9	3
Amioderone	4	1
Digoxin	1	0
Potassium	1	0
Anxiety/Depression medications	5	2
Respiratory medications	5	4

Abbreviations: SD, standard deviation; BMI, body mass index; NYHA; New York Heart Association; EF, ejection fraction; TTE; transthoracic echocardiogram, ARA 2, angiotension II receptor agonists.

*Average BMI is 18.5–24.9kg/m2 , overweight is 25–29.9kg/m2 , obese is 30kg/m2 [20]

** Two participants had missing height data affecting two in the inter-rater and one in the intra-rater calculation

*** Five participants had nil EF documented on TTE in the inter-rater group and four in the intra-rater group.

Table 2: Results for inter-rater and intra-rater reliability- ICC and Bland and Altman tests

	ICC		Bland and Altman	
	ICC co-efficient	95% CI	mean difference, $\bar{d}$ (kg)	limits of agreement (kg)
Inter-rater reliability	0.93	0.83 – 0.97	-3.81	-23.39 – 15.77
Intra-rater reliability	0.96	0.81 – 1.00	4.5	-8.11 – 17.11