J Rehabil Med 2017; 49: 689-699

REVIEW ARTICLE

Check for updates

MEASUREMENT PROPERTIES OF MAXIMAL CARDIOPULMONARY EXERCISE TEST PROTOCOLS IN PERSONS AFTER STROKE: A SYSTEMATIC REVIEW

Harriet WITTINK, PhD[1], Olaf VERSCHUREN, PhD[2], Caroline TERWEE, PhD[3], Janke DE GROOT, PhD[1], Gert KWAKKEL[4], PhD and Ingrid VAN DE PORT, PhD[5]

From the [1]Lifestyle and Health Research Group, Faculty of Health Care, Utrecht University of Applied Sciences, [2]Center of Excellence in Rehabilitation Medicine, Brain Center Rudolf Magnus, University Medical Center Utrecht, and De Hoogstraat Rehabilitation, Utrecht, [3]Department of Epidemiology and Biostatistics and the EMGO+ Institute for Health and Care Research, [4]Department of Rehabilitation Medicine, MOVE Research Institute Amsterdam, VU University Medical Center, Amsterdam, and Department of Neurorehabilitation, Reade Center for Rehabilitation and Rheumatology, Amsterdam, and [5]Revant Rehabilitation Centre Breda, Breda, The Netherlands

Objective: **To systematically review and critically appraise the literature on measurement properties of cardiopulmonary exercise test protocols for measuring aerobic capacity, VO_{2max}, in persons after stroke.**

Data sources: **PubMed, Embase and Cinahl were searched from inception up to 15 June 2016. A total of 9 studies were identified reporting on 9 different cardiopulmonary exercise test protocols.**

Study selection: **VO_{2max} measured with cardiopulmonary exercise test and open spirometry was the construct of interest. The target population was adult persons after stroke. We included all studies that evaluated reliability, measurement error, criterion validity, content validity, hypothesis testing and/or responsiveness of cardiopulmonary exercise test protocols.**

Data extraction: **Two researchers independently screened the literature, assessed methodological quality using the COnsensus-based Standards for the selection of health Measurement INstruments checklist and extracted data on measurement properties of cardiopulmonary exercise test protocols.**

Data synthesis: **Most studies reported on only one measurement property. Best-evidence synthesis was derived taking into account the methodological quality of the studies, the results and the consistency of the results.**

Conclusion: **No judgement could be made on which protocol is "best" for measuring VO_{2max} in persons after stroke due to lack of high-quality studies on the measurement properties of the cardiopulmonary exercise test.**

Key words: stroke; maximal cardiopulmonary exercise test; VO_{2max}; measurement properties.

Accepted Jun 2, 2017; Epub ahead of print Sep 1, 2017

J Rehabil Med 2017; 49: 689–699

Correspondence address: Harriet Wittink, Utrecht University of Applied Sciences Chair Lifestyle and Health Research Group Heidelberglaan 7, NL-3584 CS, Utrecht, The Netherlands. E-mail: harriet.wittink@hu.nl

Cardiopulmonary capacity (i.e. aerobic capacity or VO_{2max}) is defined as the highest rate at which oxygen can be taken up and consumed by the body during intense exercise (1). Obtaining a valid measure of aerobic capacity in people after stroke is important for the purpose of determining exercise capacity, training prescription, treatment efficacy evaluation, and/or investigation of exercise-induced adaptations of the oxygen transport/utilization system (2). The current gold standard for the assessment of aerobic capacity is considered to be the maximal cardiopulmonary exercise test (CPET) with measurements of ventilation and gas exchange, for direct assessment of maximal oxygen uptake (VO_{2max}). VO_{2max} is the accepted indicator of aerobic capacity and reflects the limits of the cardiorespiratory system to respond to exercise. Peak oxygen uptake (VO_{2peak}) reflects the highest amount of oxygen consumption attained during a test, but does not necessarily define the highest value *attainable* by the subject (3). Whereas VO_{2peak} is probably a valid index for VO_{2max} in healthy subjects (4), there is no evidence that this is the case in persons after stroke. The assessment of aerobic capacity in persons after stroke seems more challenging than in healthy subjects because they present with stroke-specific impairments, such as muscle weakness, fatigue, poor balance, contractures and spasticity, which can compromise CPET (5, 6). Marzolini et al. (7), for instance, reported that, at the start of an exercise training intervention, only 68.4% of CPETs ($n=98$) provided sufficient information to prescribe exercise intensity in persons with chronic stroke, suggesting that many persons after stroke do not reach the limits of their cardiopulmonary system before training. In the search for a CPET protocol that allows persons after stroke to reach the limits of their cardiopulmonary system, a multitude of different protocols have been developed using treadmill exercise (6, 8), body-weight supported treadmill exercise (9), (recumbent) leg cycle ergometry (6, 10, 11) and recumbent stepper exercise (12).

However, little is known about the measurement properties of these different CPET protocols in persons after stroke. To guide further research on CPET protocols in persons after stroke and to be able to interpret changes in aerobic capacity after exercise interventions, the aim of this systematic review was to critically appraise the measurement properties of

This is an open access article under the CC BY-NC license. www.medicaljournals.se/jrm
Journal Compilation © 2017 Foundation of Rehabilitation Information. ISSN 1650-1977 doi: 10.2340/16501977-2260

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

CPET protocols for measuring VO_{2max} in persons after stroke and, if possible, make recommendations for the "best" CPET protocol to use.

METHODS

Search strategy and selection criteria

A systematic review was performed in accordance with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology (13). The findings were reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (14). PubMed, Embase and CINAHL databases were searched from inception until 15 June 2016, without language or date restrictions, in order to find all available studies on measurement properties of CPET protocols in persons after stroke.

The search was performed using the medical subject heading (MESH) terms and text words (or synonyms) for the construct of interest (maximal aerobic capacity OR VO_{2max}) AND population of interest (stroke OR cerebrovascular accident OR CVA OR brain ischemia). For measurement properties, a validated search filter for finding studies on measurement properties (15) was applied in PubMed, which was adapted for use in CINAHL and Embase. The full search is detailed in Appendix SI[1]. In addition, bibliographies of included articles were searched manually for additional references. For "grey literature" the first 10 pages of Google Scholar were searched using the terms "maximal exercise test" and "cerebrovascular accident".

Study selection

The inclusion criteria were as follows: for the construct of interest and type of measurement: studies that directly measured VO_{2max} during a CPET protocol using gas exchange by open-circuit spirometry, rather than indirect estimations, that evaluated at least one measurement property of a CPET protocol. Population: persons older than 18 years after stroke. Stroke was defined according to the World Health Organization (WHO) as "a syndrome of rapidly developing symptoms and signs of focal, and at times global, loss of cerebral function lasting more than 24 h or leading to death, with no apparent cause other than that of vascular origin" (16). Finally, only full-text original articles were included

Studies were excluded when a submaximal protocol was used to predict VO_{2max} or when the only criterion for maximal testing was based on predicted maximal heart rate or a percentage of predicted maximal heart rate. Predicted heart rate was excluded since maximal heart rate response to exercise possesses wide variability (±12 beats per min) in the general population, which negatively impacts the ability to gauge subject effort by their heart rate response alone. The widespread use of beta-blocking agents in persons after stroke further complicates this issue by significantly blunting the maximal heart rate response in a disparate manner, negating the validity of the age-predicted maximal heart rate equation (17).

Eligibility criteria were applied independently by 2 reviewers (HW, IvdP) to screening titles and abstracts from all retrieved studies. Full-text review against inclusion criteria was performed of potentially eligible studies by the same 2 independent reviewers. Disagreement was resolved by joint review of the studies to reach consensus. Authors were contacted when relevant information was not available from the article.

Assessment of methodological quality

Two reviewers (HW and OV) independently evaluated the methodological quality of the included studies using the COSMIN checklist (18). The COSMIN checklist is a standardized modular tool for evaluating the methodological quality of studies on measurement properties of health-related patient-reported outcomes. However, the checklist can also be used to evaluate the methodological quality of studies on other outcome measures (18).

The COSMIN checklist contains 12 boxes: 9 of which can be used to assess the methodological quality of studies on measurement properties, 2 contain general requirements (item response theory (IRT) methods and generalizability) and 1 assesses interpretability. Interpretability is not considered a measurement property, but it is considered an important requirement for the suitability of an instrument in research or clinical practice (18). Three measurement properties (internal consistency, structural validity, cross-cultural validity) were considered not relevant for CPET protocols. Six boxes on methodological quality were used to assess the quality of the studies regarding the CPET measurement properties: reliability, measurement error, content validity, criterion validity, hypothesis testing and responsiveness.

Measurement error of CPET protocols can be established by calculating standard error of measurement (SEM). The SEM represents the extent to which CPET measurements vary if CPET is repeated without any underlying change in the patient (19). The SEM can be converted into the smallest detectable change for an individual (SDCind), which reflects the smallest within-person change in score that, with 95% confidence, can be interpreted as a ''real" change, above measurement error, in one individual. For evaluative purposes, the SDC should be smaller than the minimal amount of change in VO_{2max} that is considered to be clinically important (minimal important change; MIC) for an individual patient (20).

Content validity in this study was defined as the degree to which the result of a CPET protocol reflects VO_{2max}. We therefore assessed whether a clear description of the measurement aim, the target population, the construct to be measured (VO_{2max}) by the primary criterion (plateau VO_2) and secondary criteria (respiratory exchange ratio (RER), age predicted maximum heart rate (APMHR)) and the percentage of persons achieving these criteria, were provided. The methodological quality of a study was downgraded if duration of CPET was not reported, as the recommendation is to use protocols with modest, equal increments in work rates to achieve an exercise duration of 8–12 min (17, 21).

The items of each box were rated with a 4-point scoring system; excellent, good, fair, and poor. In line with the COSMIN checklist guidelines, an overall score for the methodological quality was obtained by taking the lowest rating of any item in a box ("worst score counts") (22). The methodological quality of a study was evaluated per measurement property.

The "Generalizability" box was used to evaluate general requirements: (*i*) whether the studies adequately described their samples in terms of age, sex, disease characteristics, setting, country, and language; (*ii*) whether they used adequate selection procedures; and (*iii*) whether acceptable missing response rates were applied. Missing response rates were defined in this study as the percentage of persons who were unable to complete the CPET protocol after inclusion. This box was used to generate a table on the characteristics of the study populations.

[1]https://doi.org/10.2340/16501977-2260

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

Data synthesis and analysis

To assess whether the results of the measurement properties were positive, negative, or indeterminate, we applied quality criteria for good measurement properties (Table I) (20). One rater (HW) applied the quality criteria.

A "best-evidence" synthesis was performed to rate the quality of the CPET protocols using the criteria in Table II. The possible levels of evidence for a measurement property are "strong", "moderate", "limited", "conflicting" or "unknown". Best-evidence synthesis was derived taking into account the methodological quality of the studies (COSMIN score; excellent, good, fair, and poor), the results (positive, indeterminate, negative) and the consistency of the results (Table II). Measurement properties from studies that were rated as poor quality were not included the best-evidence synthesis (23).

Statistical analysis

To quantify inter-rater agreement concerning the COSMIN methodological quality assessment, a linear weighted Cohen's kappa (κ) was calculated on the original items of each COSMIN box scores of the 2 reviewers (before discussion). Any disagreements in scoring were resolved after discussion.

The following interpretation was used: 0.01–0.20 means slight agreement; 0.21–0.40 means fair agreement; 0.41–0.60 means moderate agreement; 0.61–0.80 means substantial agreement; and 0.81–0.99 means almost perfect agreement (24). Where possible, SEM, and SDCindividual were calculated from reported test-retest standard deviations (SDs) and intraclass correlation coefficients (ICCs).

Table II. Best-evidence synthesis. Adapted from Terwee et al. (20)

Level	Criteria
Strong	Consistent findings in multiple studies of good methodological quality OR in 1 study of excellent methodological quality
Moderate	Consistent findings in multiple studies of fair methodological quality OR in 1 study of good methodological quality
Limited	One study of fair methodological quality
Conflicting	Conflicting findings
Unknown	Only studies of poor methodological quality

Standard error of measurement (SEM) was calculated using the following formula: $SEMagreement \sigma\sqrt{(1-ICC_{2,1})}$, where σ is the pooled standard deviation of test and retest scores. SEM can be calculated as SEMagreement or SEMconsistency (20). SEMagreement takes the systematic difference between test and retest into account, while the SEMconsistency ignores systematic differences (20). We therefore used SEMagreement in all calculations. The SEM can be converted into the smallest detectable change (SDC) (SDC = 1.96*√ 2 *SEM), which reflects the smallest within- person change in score that, with 95% confidence, can be interpreted as a ''real'' change, above measurement error, in one individual (SDCind).

The COSMIN scoring system used in this review was initially developed for assessing the quality of research into measurement properties of patient-reported outcomes (PROMs) and defines a minimum adequate sample size as 30 (fair) and adequate sample size as 100 (excellent). In this review we expected high reliabilities, because in persons with cardiac or respiratory disease the test-retest reliability was shown to be ICC = 0.95 (25–27). To determine whether sample size was

Table I. Quality criteria for measurement properties. Adapted from Terwee et al. (20)

Measurement property	Definition	Quality
Reliability	The proportion of the total variance in the measurements which is due to "true" differences between patients.	+ ICC/weighted kappa ≥ 0.70 OR Pearson's r ≥ 0.80 ? Neither ICC/weighted kappa, nor Pearson's r determined – ICC/weighted kappa < 0.70 OR Pearson's r < 0.80
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured, expressed as standard error of measurement (SEM). The SEM can be converted into the smallest detectable change (SDC). Changes smaller than the SDC can be considered measurement error and changes larger than SDC represent real change (18).	+ MIC > SDC OR MIC outside the LOA ? MIC not determined – MIC ≤ SDC OR MIC equals or inside LOA, despite adequate design and method
Content validity	The degree to which the results of CPET protocols reflect VO_{2max}. Here we determined which percentage of subjects reached pre-set criteria for maximal effort, i.e. a plateau in VO_2 or <1.5 ml/kg^{-1}/min^{-1} increase in VO_2 following workload increases, RER ≥ 1.0, failure of HR to increase with further increases in exercise intensity were reached. Duration of CPET protocol is between 8 and 12 min.	+ A clear description is provided of the measurement aim, the target population, the concept(s) being measured (VO_{2max}). More than 85% of subjects meet criteria for VO_{2max} ? A clear description is provided of the measurement aim, the target population, the concept(s) being measured (VO_{2max}). 50%–85% of subjects meet pre-set criteria for VO_{2max} - No clear description is provided of the measurement aim, the target population, the concept(s) being measured, or less than 50% of subjects meets pre-set criteria for VO_{2max}
Criterion validity	The degree to which the scores of an instrument are an adequate reflection of the gold standard protocol for (CPET), which, in healthy persons, involves an upright bicycle or treadmill protocol.	+ Convincing arguments that gold standard is ''gold'' AND correlation with gold standard > 0.70 ? No convincing arguments that gold standard is ''gold'' OR doubtful design or method – Convincing arguments that gold standard is "gold" AND Correlation with gold standard < 0.70, despite adequate design and method
Responsiveness	The ability of an instrument to detect change over time in the construct to be measured and is assessed by testing pre-specified hypotheses about the relationship between the change scores of the instrument and changes in other measures. For CPET we determined that responsiveness could only be established if at least 2 different CPET protocols were compared over time.	+ Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70 AND correlation with related constructs is higher than with unrelated constructs ? Solely correlations determined with unrelated constructs - Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlation with related constructs is lower than with unrelated constructs

+: positive rating; ?: indeterminate rating; –: negative rating; ICC: intraclass correlation coefficient; AUC: area under the curve; LOA: limits of agreement; CPET: cardiopulmonary exercise test; MIC: minimal important change; SDC: smallest detectable change.

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

adequate for calculating ICC values in the test-retest trials, we ran a simulation with 100,000 replications under the assumption that "true" ICC values would be approximately 0.95. We wanted to determine the sample size large enough to detect observed ICC values larger than 0.90 with 80% probability. The simulation we ran resulted in a required sample size of n=5 to achieve an 82% probability of measuring an ICC value >0.90.

This is assuming that the data are distributed normally. This assumption might not be met with such a small sample size. However, we decided to accept larger sample sizes than $n=5$ as adequate ("fair") for COSMIN scoring purposes. According to our simulation a sample size of $n=10$ results in a 95% probability of measuring an ICC >0.90, which we considered good, and a sample size of $n=15$ yielded a 99% probability of measuring an ICC>0.90, which we considered excellent.

RESULTS

Search

The selection procedures are summarized in Fig. 1. The search in Medline, Embase and Cinahl resulted in 1,644 citations. After removing duplicates ($n=472$), 1,172 titles and abstracts were screened by 2 reviewers independent of each other for inclusion criteria, after which 1150 citations were excluded. The full texts of the remaining 22 studies were examined in detail by the 2 reviewers. One study was retrieved from reference lists (28), increasing the total to 23 included studies. Of these 23 studies 14 did not meet the inclusion criteria; 1 was an abstract only (29); 8 studies reported on a submaximal test or used the attainment of (a percentage of) maximal predicted heart rate as their only criterion (28, 30–36), and in 3 studies no measurement properties were reported (37–39), 1 study included the same sample as an earlier (included) study (9) and 1 study did not report criteria for determining VO_{2max} (40). Searching the grey literature through the first 10 pages of Google Scholar using "maximal exercise test" and "stroke" yielded no additional articles or additional information.

Finally, 9 studies met all inclusion criteria and were included for review. The 9 studies included 2 studies with a test-retest design (6, 41), 1 RCT (10), one prospective cohort design (5), 3 cross-sectional feasibility studies (11, 42, 43), one retrospective study (7) and one within-subject design study (12).

The 9 studies reported on 9 different CPET protocols; 2 upright bicycle protocols with workload increments of 10 W/min (10) or 20 W/min (6) in persons with chronic stroke, 1 upright bicycle protocol with increments of 5 W/min in persons with acute stroke (42), 1 semi-recumbent bicycle protocol with increments of 5 W/min in persons with acute stroke (11), 2 treadmill protocols, both based on self-selected speed with different % inclines/min; 1 with body weight support (5) and 1 without (41), 1 robotics assisted tilt table (43), 1 recumbent stepper (12) and in 1 study a combination of protocols was used (7) in persons with subacute or chronic stroke.

Test-retest reliability was tested in 5 studies (5, 6, 10, 11, 41). Three studies tested test-retest reliability in a subsample of their persons who were part of a larger study (5, 10, 11). One of the 5 studies assessed measurement error (6) and in 1 study measurement error could be calculated from the available data (11). No studies explicitly mentioned content validity, but 6 studies reported on the feasibility of the test protocol (5, 7, 11, 41–43), which provides some information on content validity. In one study criterion validity was assessed (12). No studies were found in which responsiveness was measured or where hypothesis testing was performed.

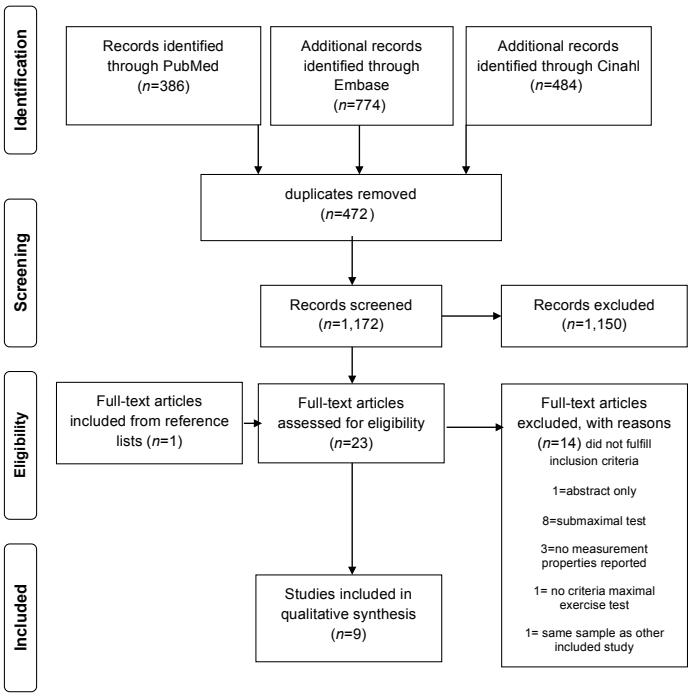

Fig. 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

Study characteristics

Study characteristics are described in Table III. Out of the 9 studies, 6 included persons with chronic stroke and 3 included persons early after stroke (mean days post-stroke 9.9 (standard deviation (SD) 2.0) (42), 17.6 (SD 2.2) (11) and 26.0 (SD 8.8) (5)). Sample sizes ranged from 6 (5) to 98 (7) persons. All studies, except one (10) specified where testing took place. Six studies described the method to select their subjects (random, consecutive or convenience), in 3 studies this was unclear (5, 41, 43).

Table III. General characteristics of included studies

Author, year (reference)	Country	Sample size *n*	Age Mean (SD)	Males %	Mean stroke severity	Stroke duration	Beta-blockers *n*/%	Protocol	Method selection subjects	Measurement property
Potempa et al., 1995 (10)	US	25	NR	NR	NR	>6 months	NR	Bicycle ergometer	Convenience sample of rehabilitation completers	Test-retest reliability
Mackay-Lyons & Makrides 2002 (5)	Canada	6	NR	67	CMSA leg 5.3 (0.7)	>2 months	NR	Treadmill + 15% BWS	from acute stroke service	Test-retest reliability
		29	64.9 (13.5)	76			15/52			Content validity
Dobrovolny et al., 2003 (41)	US	53	64 (8)	83	Mild-to-moderate gait deficits	>6 months	NR	Treadmill	Physician referrals in rehabilitation hospital	Test-retest reliability
										Content validity
Eng et al. 2004 (6)	Canada	12	62.5 (8.6)	92	CMSA leg and foot = 9.4 (2.5)	3.5 (2.0) years	NR	Bicycle ergometer	Recruitment community dwelling adults on volunteer basis	Test-retest reliability
Tang et al., 2006 (11)	Canada	20	69.3 (2.3)	60	CMSA leg = 4.7 (0.2)	18.6 (3.1) days	0/0	Semi-recumbent bicycle	Consecutive sample rehabilitation facility	Test-retest reliability Measurement error
		35	65.7 (2.6)	54	CMSA leg = 4.7 (0.2)	17.6 (2.2) days	5/14			Content validity
Chen et al., 2010 (42)	Taiwan	19	62.7 (9.2)	100	MAS = 0.2 (0.4)	9.9 (2.0) days	8/42 Calcium-channel blockers	Bicycle ergometer	Convenience sample acute stroke service	Content validity
Saengsuwan et al., 2015 (43)	Switzerland	8	58.3 (9.2)	50	FAC = 1.8	>1 year	1/12.5	Augmented robotics assisted tilt table	? Rehabilitation centre	Content validity
Marzolini et al., 2012 (7)	Canada	98	63.3 (12)	68	CMSA leg = 5.0 (1.1)	≥3 months	42/43	Recumbent bike, upright bike, treadmill	Consecutive sample rehabilitation facility	Content validity
Billinger et al., 2008 (12)	US	11	60.9 (12)	64	LEFM = 25.7 (6.4)	40.1 (32.7) months	9/82	Total-body recumbent stepper	?	Criterion validity

?: indeterminate rating; CMSA: Chedoke-McMaster Stroke Assessment; FM: Fugl-Meyer; LEFM: Lower Extremity Fugl-Meyer; MAS: Modified Ashworth Scale; FAC: Functional Ambulation Category; BWS: body weight support; NR: not reported.

All studies included persons with mild-to-moderate impairments, except for Billinger et al. (12), who included persons with mild-to-severe impairments according to mild-to-severe lower extremity Fugl-Myer test scores, and Shaengsuwan et al. (43) who included dependent ambulatory persons with a mean Functional Ambulation Score (FAC) score of 1.8.

Measurement properties and methodological quality

The inter-rater agreement of the independent methodological quality assessment of included studies was linear weighted kappa = 0.86 (95% confidence interval (95% CI) 0.77–0.96). This was interpreted as almost perfect agreement (24). Disagreement was mainly due to errors in interpretation and was easily resolved using a consensus method between the 2 raters.

Reliability

Among the 5 test-retest studies (total $n = 116$), 2 studies reported the evaluation of the test-retest reliability of CPET as the main purpose of their study (6, 41). In these 2 studies sample sizes ranged from 12 (6) to 53 (41) persons with chronic stroke (>6 months), with mean ages ranging from 62.5 (SD 8.6) (6) to 64 (8) years. The other 3 studies (5, 10, 11) reported on testing a subsample of their subjects within a larger study and included sample sizes ranging from $n = 6$ (5) to $n = 25$ (10) and persons with acute (5, 11) and chronic stroke (10).

Five different protocols were found, including treadmill (6, 41), body weight supported treadmill (5) and (recumbent) leg cycle ergometry (6, 10, 11).

ICC values were higher than 0.90 for 3 studies (5, 6, 10) 1 study reported an ICC of 0.50 (11) and 1 study reported an undefined correlation of 0.92 (41).

The methodological quality of the reliability studies was rated as poor (5, 11), fair (6, 41) and good (10). In all studies there was ambiguity about independent administrations. None of the studies specified whether subjects were stable in the interim period. The reported interim period was 1–4 days (11), 2 days (10), 3–4 days (5) and 1 week (41). One author did not report the interim time-period (6).

Two out of the 5 studies reported on whether test conditions were similar for both measurements (11, 41). Only 1 study explicitly reported similar test conditions using the same equipment and raters for both tests (11). The main methodological flaws included inappropriate or not-reported time intervals between tests (6, 11) and selection bias (5, 11).

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

Best-evidence synthesis reliability

Due to poor methodological quality 2 of the 5 studies were excluded from the best-evidence synthesis (5, 11). In the 3 remaining studies 3 different protocols were used; 2 different upright bicycle protocols (6, 10), and a treadmill protocol (41). As the 2 bicycle protocols used similar pedalling rates and increases in Watts per min, we felt we could combine the results in an evidence synthesis, and concluded there is moderate evidence for positive reliability in upright bicycle protocols in persons with chronic stroke. In addition, there is limited evidence for positive reliability for a treadmill protocol (41), based on one study of fair quality (Table IV).

Table IV. Methodological quality/measurement properties and best evidence

Author	Sample	Protocol	Reliability	Measurement error	Content validity	Criterion validity
Upright bicycle						
Potempa et al., 1995 (10)	Chronic stroke	Upright bicycle: seated rest on the bicycle ergometer for 2 min. Dynamic exercise began at 10 W, and workloads increased by 10 W each min. Pedalling rate was between 50 and 70 rpm	Good/positive			
Eng et al., 2004 (6)	Chronic stroke	Upright bicycle Subjects began pedalling at 0 W with workload increments of 20 W/min and were instructed to pedal at a comfortable rate which was generally between 50 and 70 rpm	Fair/positive	Fair/negative		
		Best evidence	Moderate/+	Limited/–	?	?
Upright bicycle						
Chen et al., 2010 (42)	Acute post stroke 9.9±2.0 days	5-W ramp-incremental protocol for stroke patients to either voluntary exhaustion or the inability to maintain a pedalling rate of 50 rpm. The subjects were required to maintain a constant pedalling rate at 60 rpm at all power outputs			Good/indeterminate	
		Best evidence	?	?	?	?
Semi-recumbent bicycle						
Tang et al., 2006 (11)	acute post stroke, 17.6±2.2days	The ramp protocol included a 2-min warm-up at 10 W at a target cadence of 50 rpm, followed by progressive 5 W increases in work rate every min	Poor/negative	Poor/negative	Good/indeterminate	
		Best evidence	?	?	?	?
Treadmill						
Dobrovolny et al., 2003 (41)	Chronic stroke	Macko protocol (8)/treadmill determined walking speed During the initial 2 min, subjects walked on the treadmill without an incline, followed by 2 min at 4% incline, with the incline advanced 2% per 2 min thereafter.	Fair/positive		Good/negative	
		Best evidence	Limited/+	?	Moderate/–	?
Treadmill + weight support						
Mackay-Lyons et al., 2002 (5)	Post stroke <1 month	Treadmill + weight support Stages 2 min each with 15% of body mass suspended Warm-up 1min at 60–70% self-selected speed and no incline	Poor/positive		Good/indeterminate	
		Stage 1 Self-selected treadmill speed with no incline Stages 2–5 Self-selected treadmill speed maintained and 2.5% increase in incline at each stage Stages >6 10% incline maintained and 0.05 m/s increase in treadmill speed at each stage				
		Cool-down 2 min at 60–70% self-selected speed and no incline				
		Best evidence	?	?	?	?
Robotics-assisted tilt table (RATT)						
Saengsuwan et al., 2015 (43)	Chronic stroke Unable to walk independently	1. Rest phase for 3 min; 2. A passive phase for 5 min; 3. A constant-load phase, where the patient actively moved their legs in synchrony with the RATT motion to follow the constant target work rate (the work rate was set at 40 % of peak work rate (WRpeak) obtained from the IET) for 10 min; and 4. A recovery phase for 5 min.			Good/positive	
		Best evidence	?	?	Limited/+	?
Recumbent stepper						
Billinger et al., 2008 (12)	Chronic stroke	Bicycle protocol: resistance was set at 0 W for the first 3 min of the exercise test and then was increased by 10 W/min until test termination. Modified Total-Body Recumbent Stepper Exercise Test: Start with 25 W, Increase 15 W every 2 min while maintaining 80 rpm.				Fair/positive
		Best evidence	?	?	?	Limited/+
Combination						
Marzolini et al., 2012 (7)	Chronic stroke	A CPET on either a recumbent cycle ergometer with specialized pedals to secure feet, an upright cycle, or a treadmill was performed. The type of ergometer and testing protocol was chosen by the testing staff based on balance and control of leg/foot position on the pedal			Good/negative	
		Best evidence	?	?	Moderate/–	?

?: indeterminate rating; CPET: cardiopulmonary exercise test; IET: Institution of Engineering and Technology.

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

Measurement error

Of the 5 reliability studies only one reported on measurement error in terms of standard error of measurement (SEM) (6). SDCindividual was calculated from the SEM given by Eng et al. (6), although we were not able to verify whether SEM was calculated correctly, as test-retest data were not given in the paper. We were able to calculate SEM and SDCindividual for one more study based on the pooled SD of test and retest SD (11). Mean SEM varied between 0.36 ml/kg/min (11) and 1.0 ml/kg/min (6). SDCindividual ranged from 1.0 ml/kg/min in persons with sub-acute stroke (11) to 2.77 ml/kg/min in persons with chronic stroke (6).

One study was rated of fair quality (6) due to not reporting of the time interval between measurements. Four studies were rated as poor due to not calculating any measurement error, as we would expect in a reliability study.

Best-evidence synthesis for measurement error

As we could not find any research on minimal clinically important change (MIC) in persons after stroke, we set the MIC for oxygen uptake at 1 ml/kg/min based on the study by Keteyian et al. (44), which found that in persons with coronary heart disease VO_{2peak} is a strong predictor of all-cause death, with every 1 ml/kg/min increase in VO_{2peak} associated with an approximate 15% decrease in risk of death. The MIC was smaller than the SDCindividual in one study of fair quality (6) and equal to the SDCindividual in one poor-quality study (11). Excluding the poor-quality study from the best-evidence synthesis, there is limited evidence for negative measurement error in an upright bicycle protocol (Table IV).

Content validity

Out of the 9 studies, 6 reported the criteria for determining the attainment of VO_{2max} and the percentage of persons that reached these criteria (5, 7, 11, 41–43) (Table V). None of these studies set out to explicitly test content validity, but they set out specifically to determine the feasibility of a protocol.

Criteria for determining VO_{2max} were not consistent across studies (see Table V).

Two studies did not report on the formula used for determining APMHR (11, 41) and 2 studies used an adjusted formula for subjects taking beta-blockers (5, 7). Only one study reported on the duration of testing (5).

The quality of all 6 studies was rated as good. In these studies consistently a clear description was provided of the measurement aim, the target population and the concept(s) being measured (VO_{2max}). Subject achievement of pre-set criteria for VO_{2max} was 87.5% in 1 study (43), 50–80% in 3 studies (5, 11, 42) and less than 50% in 2 studies (7, 41) ($n=151$). Main methodological flaws were including systolic blood pressure >200 mmHg as a criterion for attainment of VO_{2max} (5), incomplete reporting on criteria set for determination of VO_{2max} (41) and not reporting on duration of testing (7, 11, 41–43).

Best-evidence synthesis of content validity

The studies could not be combined due to the difference between the study protocols; therefore we determined the evidence for content validity per protocol.

In 3 good-quality studies (5, 11, 42) ($n=83$) we rated the evidence for content validity as unknown for a body weight supported treadmill protocol, a semi-recumbent, and an upright bicycle protocol. In 2 good quality studies (7, 41) ($n=151$) in which less than 50% of subjects achieved criteria for VO_{2max}, we found moderate evidence for negative content validity of a treadmill protocol (8) and protocols chosen by testing staff (7).

In 1 study of good quality 87.5% of subjects met pre-set criteria for VO_{2max}. However, the sample size was small ($n=8$), we therefore downgraded the evidence to limited evidence for positive content validity of a

Table V. Criteria for construct validity cardiopulmonary exercise test (CPET) protocols

Author	Plateau VO_2	Definition plateau	RER	Heart rate	Formula APMHR	% Patients achieving criteria
Mackay-Lyons et al., 2002 (5)	Yes	VO_2 increase < 150 ml/min in last min exercise	> 1.0	HR_{peak} ± 15 bpm APMHR	220 – age or for patients taking beta-blockers: (220 – age)*85%	76
Dobrovolny et al., 2003 (41)			> 1.1	Achievement of APMHR		9
Tang et al., 2006 (11)	Yes		> 1.0	85% of APMHR		71
Chen et al., 2010 (42)	Yes	VO_2 increase < 150 ml/min final min exercise	> 1.0	HR_{peak} ± 15 bpm APMHR	220 – age	78.9
Saengsuwan et al., 2015 (43)	Yes	VO_2 increase < 150 ml/min final min exercise	> 1.1 or > 1.05 for age 50+	$HR_{peak} \geq HR_{max}$ – 10 bpm	220 – age	87.5
Marzolini et al., 2012 (7)	Yes	VO_2 increase < 2.1 ml/kg min for ≥ 60 s despite an increase in work rate	> 1.15	HR_{peak} ± 10 beats of APMHR	220 – age or for patients taking beta-blockers: 164–0.7 age	18.4

RER: respiratory exchange ratio; APMHR: age predicted maximum heart rate; bpm: beats/min.

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

robotics-assisted tilt table in dependent ambulatory persons after stroke (43) (Table IV).

Criterion validity

Only one study determined criterion validity. A semi-recumbent stepper protocol was compared with an upright bicycle ergometer protocol (gold standard protocol) (12). The methodological quality of this study for criterion validity was rated as fair due to the small sample size ($n = 11$).

Best-evidence synthesis of criterion validity

The level of evidence for criterion validity was provided by one study of fair quality (12). With a correlation of $r = 0.91$ between VO_{2peak} obtained by the 2 protocols, the criterion validity measurement was rated as positive. This study therefore provided limited evidence of positive criterion validity of the recumbent stepper protocol compared with a cycle protocol in persons with chronic stroke with mild-to-severe impairments.

DISCUSSION

The results of this review reveal a lack of high-quality studies into the measurement properties of CPET protocols for persons after stroke. Nine studies were found, which reported 9 different CPET measurement protocols. Most studies reported on only one measurement property. No studies were found on hypothesis testing or responsiveness. In most of the studies in this review a substantial proportion of subjects were reported to not reach the limits of their cardiopulmonary systems during the CPET protocol. Although authors agreed on a VO_2 plateau as a primary criterion to determine the attainment of VO_{2max}, we found substantial variation between secondary criteria. As a plateau in VO_2 is often not found, secondary criteria are frequently relied upon to determine attainment of VO_{2max}, which means that these criteria directly impact on the content validity of CPET protocols.

At present, due to the heterogeneity of CPET protocols and secondary criteria, as well as the scanty information on measurement properties of CPET protocols, we are unable to make recommendations as to which protocol to use to "best" measure VO_{2max} in persons after stroke. However, for readers interested in recommendations for clinical practice, we refer to our recent systematic review (45).

Reliability

Test-retest reliability depends on daily or weekly fluctuations in VO_{2max} and is based on the assumption that no real change in VO_{2max} occurs between tests. Subjects who have undergone maximal exercise testing will need time to recover from this exhaustive work. Although there is no research available to guide an appropriate time interval between tests in persons with stroke, we considered 1 day (11) as too short to allow these deconditioned individuals to recover from exhaustive testing. Although the CPET protocols used were diverse, and the time interval varied between 1 and 7 days, ICC values between the reliability studies were remarkably similar. There was moderate evidence for positive reliability for upright bicycle CPET protocols, consistent with studies on healthy adults and in persons with cardiac or respiratory disease (46–48). The one exception was the study by Tang et al. (11) reporting an ICC of 0.50. This study also reported the smallest SD of VO_{2peak} values, representing a fairly homogeneous sample of subjects, which probably also explains the low ICC value found in this study. This homogeneity might be the result of selection bias (taking the "best" subjects for test-retest). We found moderate evidence for positive reliability in CPET in 2 upright bicycle protocols (6, 10) in persons with chronic stroke. However, CPET with an upright bicycle protocol may not be able to distinguish between clinically relevant change and measurement error in individual persons after stroke.

Measurement error

Measurement error can occur due to technical factors, natural variation in the subject, variation in the measurement process, or a combination of these factors (1, 49).

Individual measurement error ($SDC_{individual}$) ranged from 1.0 ml/kg/min in persons with sub-acute stroke (11) to 2.77 ml/kg/min (6) in persons with chronic stroke and was equal to or larger than the minimal clinical important change (MIC) we set at 1 ml/kg/min based on the available literature. Therefore, clinically relevant changes cannot be distinguished from measurement error in individual persons with 95% confidence (50). It would thus be helpful if future studies, instead of reporting group mean improvements and statistical significance, would report on the percentage of individual persons who achieved a clinically important change in aerobic capacity.

Validity

Secondary criteria for the determination of attainment of VO_{2max} directly impact content validity. Although all studies reported a VO_2 plateau as the gold standard for attainment of VO_{2max}, secondary criteria varied substantially. For RER a range of 1.0–1.15 was used, but RER values greater than 1.1 are probablytoo high for persons after stroke as they tend to be older. For

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

instance, the 2 studies using RER > 1.1 reported 91% (41) and 61.8% (7) of their sample (mean age 60+) to not reach the limits of their cardiopulmonary systems (VO_{2max}), despite reported maximal effort. Edvardsen et al. (51) report RER to decline by age and recommend using a RER of ≥ 1.05 for 50–64-year-olds and RER ≥1.0 for those 65 years and older for both males and females. We propose to adopt these criteria.

We found only one study in which protocols were compared to determine which protocol is most valid in measuring VO_{2max} in persons after stroke. Billinger et al. (12) compared a semi-recumbent stepper protocol with an upright bicycle ergometer protocol in persons with mild to severe impairments, and found that VO_{2peak} values were significantly higher for the semi-recumbent stepper protocol, although RER values (both RER = 1.1 (0.1)) were not significantly different. Unfortunately, they did not report on the percentage of persons that attained criteria for VO_{2max} in either protocol.

Study limitations

We struggled with the definition of "content validity" as given by Terwee et al. (20), as what we wanted to know was whether actual CPET results reflected criteria set for VO_{2max}. As there are no criteria for content validity assessment of CPET, we determined our own. These criteria included the reporting of the duration of CPET as we feel this helps in interpreting CPET results. Using this criterion resulted in lower quality scores for 3 of the 6 content validity studies that did not report duration of testing (5, 7, 11). One could consider this too strict, as, formally, duration of testing is not a criterion for determining VO_{2max}.

We were unable to find what constitutes a MIC in aerobic capacity in persons after stroke. Therefore we used the findings by Keteyian et al. (44), as these were based on persons with coronary heart disease, and we felt this population resembled the stroke population. A recent systematic review (45) concluded that most CPET studies do not or insufficiently adhere to existing cardiopulmonary guidelines. Of the 9 included studies 8 reported adhering to ACSM guidelines for contraindications to maximal exercise testing and for termination criteria. Only Billinger et al. (29) did not refer to any guideline. Most studies did not report on pre-exercise testing or abstinence from alcohol, caffeine or tobacco prior to testing. This may have affected the reproducibility and validity of the CPET protocols.

Although we performed a comprehensive search for studies on measurement properties of CPET protocols, we may have missed studies, especially sub-studies on measurement properties embedded in larger (clinical) studies.

Conclusion

This review reveals the lack of high-quality measurement property studies on CPET protocols for persons after stroke. Our findings show the urgent need for further measurement property studies on CPET protocols in persons after stroke. We need to find consensus on which secondary criteria to use in CPET protocols in persons after stroke to determine attainment of VO_{2max}. Given the reported difficulty of measuring VO_{2max} in persons after stroke, we may need to consider an alternative; for instance, the determination of the ventilatory anaerobic threshold (VAT), or the respiratory compensation point (RCP). The VAT and RCP are recommended as an appropriate target intensity level for the prescription of exercise as they are effort-independent measures and maximum testing is not necessary. Measuring VAT in persons after stroke is feasible for the majority of subjects, as evidenced by the reports of 5 studies (7, 11, 42, 43, 52). A recent study has found the determination of VAT to have good reliability (ICC3,2 = 0.87, 95% CI 0.80–0.95) in persons after stroke (52). RCP has been shown to be identifiable in 96% of persons after stroke and to have reasonably good reliability (ICC3,1 = 0.77 (95% CI 0.24, 0.87)) (53).

Further high-quality research is needed into the measurement properties of different CPET protocols in the acute and, more severely impaired, chronic population of persons with stroke.

ACKNOWLEDGEMENTS

This study was funded by Utrecht University of Applied Sciences. The authors would like to thank Jurgen Mollema MSc for his support with constructing the search string and searching the databases.

The authors declare no conflicts of interest.

REFERENCES

1. Howley ET, Bassett DR, Jr., Welch HG. Criteria for maximal oxygen uptake: review and commentary. Med Sci Sports Exerc 1995; 27: 1292–1301.
2. Mezzani A, Agostoni P, Cohen-Solal A, Corra U, Jegier A, Kouidi E, et al. Standards for the use of cardiopulmonary exercise testing for the functional evaluation of cardiac patients: a report from the Exercise Physiology Section of the European Association for Cardiovascular Prevention and Rehabilitation. Eur J Cardiovasc Prev Rehabil 2009; 16: 249–267.
3. Whipp BJ. The peak versus maximum oxygen uptake issue. CPX international 2010. (4) Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ. The maximally attainable VO2 during exercise in humans: the peak vs. maximum issue. J Appl Physiol 2003; 95: 1901–1907.
4. Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ. The maximally attainable VO2 during exercise in humans: the peak vs. maximum issue. J Appl Physiol 2003; 95: 1901–1907.
5. Mackay-Lyons MJ, Makrides L. Exercise capacity early after

stroke. Arch Phys Med Rehabil 2002; 83: 1697–1702.
6. Eng JJ, Dawson AS, Chu KS. Submaximal exercise in persons with stroke: test-retest reliability and concurrent validity with maximal oxygen consumption. Arch Phys Med Rehabil 2004; 85: 113–118.
7. Marzolini S, Oh P, McIlroy W, Brooks D. The feasibility of cardiopulmonary exercise testing for prescribing exercise to people after stroke. Stroke 2012; 43: 1075–1081.
8. Macko RF, Katzel LI, Yataco A, Tretter LD, DeSouza CA, Dengel DR, et al. Low- velocity graded treadmill stress testing in hemiparetic stroke patients. Stroke 1997; 28: 988–992.
9. Mackay-Lyons MJ, Makrides L. Longitudinal changes in exercise capacity after stroke. Arch Phys Med Rehabil 2004; 85: 1608–1612.
10. Potempa K, Lopez M, Braun LT, Szidon JP, Fogg L, Tincknell T. Physiological outcomes of aerobic exercise training in hemiparetic stroke patients. Stroke 1995; 26: 101–105.
11. Tang A, Sibley KM, Thomas SG, McIlroy WE, Brooks D. Maximal exercise test results in subacute stroke. Arch Phys Med Rehabil 2006; 87: 1100–1105.
12. Billinger SA, Tseng BY, Kluding PM. Modified total-body recumbent stepper exercise test for assessing peak oxygen consumption in people with chronic stroke. Phys Ther 2008; 88: 1188–1195.
13. COSMIN. VU University Medical Center 2016. [Accessed 2017 apr 12]. Available from: www.cosmin.nl.
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010; 8: 336–341.
15. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009; 18: 1115–1123.
16. World Health Organization. Recommendations on stroke prevention, diagnostics and therapy: report of the WHO task force on stroke and other cerebrovascular disorders. Stroke 1989; 20: 1407–1431.
17. Balady GJ, Arena R, Sietsema K, Myers J, Coke L, Fletcher GF, et al. Clinician's guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. Circulation 2010; 122: 191–225.
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010; 19: 539–549.
19. Harvill LM. Standard error of measurement. Educational Measurement: Issues and Practice 1991; 10: 33–41.
20. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34–42.
21. Fletcher GF, Balady GJ, Amsterdam EA, Chaitman B, Eckel R, Fleg J, et al. Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association. Circulation 2001; 104: 1694–1740.
22. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012; 21: 651–657.
23. de Vet H, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. Cambridge: Cambridge University Press; 2011.
24. Cohen J. "Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit". Psychol Bull 1968; 70: 213–220.
25. Wasserman K, Hansen JE, Sue DY, Stringer WW, Whipp BJ. Principles of exercise testing and interpretation. 4th ed. Philadelphia: Lippincott Williams and Wilkins; 2005.
26. Arena R, Myers J, Williams MA, Gulati M, Kligfield P, Balady GJ, 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: 329–343.
27. Palange P, Ward SA, Carlsen KH, Casaburi R, Gallagher CG, Gosselink R, et al. Recommendations on the use of exercise testing in clinical practice. Eur Respir J 2007; 29: 185–209.
28. Duncan P, Studenski S, Richards L, Gollub S, Lai SM, Reker D, et al. Randomized clinical trial of therapeutic exercise in subacute stroke. Stroke 2003; 34: 2173–2180.
29. Billinger SA, Mattlage AE, Lentz AA, Rippee MA. Submaximal exercise test to predict VO2 peak in subacute stroke. Stroke 2013; 44: (Suppl 1).
30. Danielsson A, Sunnerhagen KS. Oxygen consumption during treadmill walking with and without body weight support in patients with hemiparesis after stroke and in healthy subjects. Arch Phys Med Rehabil 2000; 81: 953–957.
31. Danielsson A, Willen C, Sunnerhagen KS. Measurement of energy cost by the physiological cost index in walking after stroke. Arch Phys Med Rehabil 2007; 88: 1298–1303.
32. Yates JS, Studenski S, Gollub S, Whitman R, Perera S, Lai SM, et al. Bicycle ergometry in subacute-stroke survivors: feasibility, safety, and exercise performance. J Aging Phys Act 2004; 12: 64–74.
33. Lennon OC, Denis RS, Grace N, Blake C. Feasibility, criterion validity and retest reliability of exercise testing using the Astrand-rhyming test protocol with an adaptive ergometer in stroke patients. Disabil Rehabil 2012; 34: 1149–1156.
34. Tsuji T, Liu M, Tsujiuchi K, Chino N. Bridging activity as a mode of stress testing for persons with hemiplegia. Arch Phys Med Rehabil 1999; 80: 1060–1064.
35. Stoller O, Schindelholz M, Bichsel L, Schuster C, de Bie RA, de Bruin ED, et al. Feedback-controlled robotics-assisted treadmill exercise to assess and influence aerobic capacity early after stroke: a proof-of-concept study. Disabil Rehabil Assist Technol 2013; 9: 271–278.
36. Ovando AC, Michaelsen SM, Carvalho T, Herber V. Evaluation of cardiopulmonary fitness in individuals with hemiparesis after cerebrovascular accident. Arq Bras Cardiol 2011; 96: 140–147.
37. Baert I, Daly D, Dejaeger E, Vanroy C, Vanlandewijck Y, Feys H. Evolution of cardiorespiratory fitness after stroke: a 1-year follow-up study. Influence of prestroke patients' characteristics and stroke-related factors. Arch Phys Med Rehabil 2012; 93: 669–676.
38. Kelly JO, Kilbreath SL, Davis GM, Zeman B, Raymond J. Cardiorespiratory fitness and walking ability in subacute stroke patients. Arch Phys Med Rehabil 2003; 84: 1780–1785.
39. Gjellesvik TI, Brurok B, Hoff J, Torhaug T, Helgerud J. Effect of high aerobic intensity interval treadmill walking in people with chronic stroke: a pilot study with one year follow-up. Top Stroke Rehabil 2012; 19: 353–360.
40. Olivier C, Dore J, Blanchet S, Brooks D, Richards CL, Martel G, et al. Maximal cardiorespiratory fitness testing in individuals with chronic stroke with cognitive impairment: practice test effects and test-retest reliability. Arch Phys Med Rehabil 2013; 94: 2277–2282.
41. Dobrovolny CL, Ivey FM, Rogers MA, Sorkin JD, Macko RF. Reliability of treadmill exercise testing in older patients with chronic hemiparetic stroke. Arch Phys Med Rehabil 2003; 84: 1308–1312.
42. Chen JK, Chen TW, Chen CH, Huang MH. Preliminary study of exercise capacity in post-acute stroke survivors. Kaohsiung J Med Sci 2010; 26: 175–181.
43. Saengsuwan J, Huber C, Schreiber J, Schuster-Amft C, Nef T, Hunt KJ. Feasibility of cardiopulmonary exercise testing and training using a robotics-assisted tilt table in dependent-ambulatory stroke patients. J Neuroeng Rehabil 2015; 12: 88.

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM

44. Keteyian SJ, Brawner CA, Savage PD, Ehrman JK, Schairer J, Divine G, et al. Peak aerobic capacity predicts prognosis in patients with coronary heart disease. Am Heart J 2008; 156: 292–300.
45. van de Port, I, Kwakkel G, Wittink H. Systematic review of cardiopulmonary exercise testing post stroke: Are we adhering to practice recommendations? J Rehabil Med 2016; 47: 881–900.
46. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. J Appl Physiol 1955; 8: 73–80.
47. McArdle WD, Katch FI, Pechar GS. Comparison of continuous and discontinuous treadmill and bicycle tests for max Vo2. Med Sci Sports 1973; 5: 156–160.
48. Barron A, Dhutia N, Mayet J, Hughes AD, Francis DP, Wensel R. Test-retest repeatability of cardiopulmonary exercise test variables in patients with cardiac or respiratory disease. Eur J Prev Cardiol 2014; 21: 445–453.
49. Bland JM, Altman DG. Measurement error. BMJ 1996; 313 (7059): 744.
50. Terwee CB, Terluin B, Knol DL, de Vet HC. Combining clinical relevance and statistical significance for evaluating quality of life changes in the individual patient. J Clin Epidemiol 2011; 64: 1465–1467.
51. Edvardsen E, Scient C, Hansen BH, Holme IM, Dyrstad SM, Anderssen SA. Reference values for cardiorespiratory response and fitness on the treadmill in a 20–85–year-old population. Chest 2013; 144: 241–248.
52. Bosch PR, Holzapfel S, Traustadottir T. Feasibility of measuring ventilatory threshold in adults with stroke-induced hemiparesis: implications for exercise prescription. Arch Phys Med Rehabil 2015; 96: 1779–1784.
53. Saengsuwan J, Berger L, Schuster-Amft C, Nef T, Hunt KJ. Test-retest reliability and four-week changes in cardiopulmonary fitness in stroke patients: evaluation using a robotics-assisted tilt table. BMC Neurol 2016; 16: 163.

JRM Journal of Rehabilitation Medicine JRM Journal of Rehabilitation Medicine JRM