Test-retest reliability of a maximal arm cycle exercise test for younger individuals with traumatic lower limb amputations

Christensen, Jan; Tang, Lars Hermann; Doherty, Patrick; Lagnhorn, Christian; Langberg, Henning

Published in:
European Journal of Physiotherapy

DOI:
10.1080/21679169.2019.1578824

Publication date:
2020

Document version
Accepted manuscript

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 15. May. 2020
TITLE

Test-retest reliability of a maximal arm cycle exercise test for younger individuals with traumatic lower limb amputations

Running head: Upper-body CPET for younger amputees

Article category: Original Article

Authors:

Jan Christensen123, Lars Tang34; Patrick Doherty5, Christian Langhorn1, Henning Langberg2.

1 Department of Occupational- and Physiotherapy, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

2 CopenRehab, Section of Social Medicine, Dept. of Public Health, Faculty of Health, University of Copenhagen, Copenhagen, Denmark

3 Danish Knowledge Centre for Rehabilitation and Palliative Care, University of Southern Denmark, Nyborg, Denmark

4 Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Region Zealand, 4200 Slagelse, Denmark

5 Department of Health Sciences, University of York, York, UK

Corresponding author:

Jan Christensen, ORCID ID 0000-0003-1114-746 Department of Occupational- and Physiotherapy, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9 Copenhagen, 2100, Denmark.

Email: fysjan@gmail.com

Declaration of interest:

The authors report no declarations of interest. The first author is funded by a strategic grant from the Danish Defense Agreement 2013–2017.
ABSTRACT

**Study Design:** Cross-sectional study design

**Background:** Cardiopulmonary exercise testing (CPET) using the upper-body is an alternative to standardized CPET. A variety of test protocols has been developed for CPET performed on an arm crank for individuals with lower limb disabilities. However, no test protocol specifically designed to evaluate exercise capacity in younger adults with lower limb amputees is available.

**Objective:** To determine test-retest reliability of VO2peak, watts, test time and max heart rate for an upper-body CPET.

**Methods:** Males with unilateral lower limb amputation at crus-level or above, aged 18-40 years, without comorbidities preventing arm cycling were eligible. Participants completed three maximal tests with a minimum of two days between tests and maximum 14 days between the first and third test.

**Results:** Eight participants volunteered to participate and were tested three times. Excellent and acceptable ICC values were found for power in Watts (ICC=0.73) and test time (ICC=0.74). Good but non-acceptable test-retest reliability was found for VO2peak (ICC=0.51) and max heart rate (ICC=0.64) for all three test sessions.

**Conclusions:** Upper-body CPET shows acceptable test-retest reliability for Watts and test time in younger adult lower limb amputees, but non-acceptable reliability for VO2peak and heart rate.

**Keywords:** Amputation, reproducibility of results, exercise test, oxygen

Word count article: 2591
**BACKGROUND**

Measurement of maximal aerobic exercise capacity (fitness) is a core element for accurate exercise prescription. Cardiopulmonary exercise testing (CPET) performed by the upper-body using an arm ergometer or arm crank is feasible and established as an alternative to standardized CPET lower-limb protocols performed on a treadmill or on a stationary bicycle ergometer (1). Good concurrent validity has been found between CPET performed with the upper-body and CPET performed with the lower limb, if adjusted for the amount of involved muscle mass (2). Furthermore, CPET performed on an arm crank is predictive for walking ability (3, 4) and prosthesis fitting (5) in elderly vascular amputees.

Perfect achievement of all psychometric properties for a test is rare but coming close to meeting acceptable standards for clinical practice and research is essential. Low test-retest reliability of a measurement of fitness can lead to imprecise exercise training prescriptions and difficulties when interpreting a potential intervention effect (6). Furthermore, systematic measurement error would lead to additional difficulties in interpreting test results and the effect of an intervention.

A variety of test protocols have been developed for CPET performed on an arm crank for individuals with lower limb disabilities. Recommendations state that specific protocols should be developed and tested for specific groups. This may explain why a range of current protocols exist, including: Wingate test protocols (7), submaximal aerobic test protocols (8, 9) and maximal aerobic test protocols (9-11), primarily developed and tested in individuals with varying lower limb disabilities (e.g. elderly individuals with amputations or stroke).

To our knowledge an upper-body CPET protocol specifically designed for testing aerobic capacity of younger adults with lower limb amputation has not been tested in terms of reliability. The
The objective of this study was therefore to determine the test-retest reliability of a cardiopulmonary exercise test specifically designed to evaluate exercise capacity in younger lower limb amputees.

MATERIALS AND METHODS

Participants

Eligible for this study were males who had been hospitalised at the [Name of hospital removed for blinded peer-review process] after a unilateral lower limb amputation at crus-level or above, aged 18-40 years, without comorbidities precluding arm cycling. Participants were excluded if they had previous specific training on an arm crank. A purposive sampling method was used to identify eligible participants (12). Participation was voluntary and the participants received oral and written information about the study prior to giving written informed consent to participate in the study. The study was approved by the local ethics committee (H-15020067) and the handling of data was approved by the local Data Protection Agency (j.nr.: 2012-58-0004). The participants received financial compensation for transportation in relation to the study.

Data collection

Each participant completed three test sessions between October 2016 and February 2017. All tests were performed at the same time of day with a minimum of two days between tests, and a maximum of 14 days between the first and the third test. Before the first test session, written informed consent was obtained, along with participants’ characteristics, including demographic data, resting systolic and diastolic blood pressure (BP), resting heart rate, resting subjective perceived exertion using the 15-point Borg scale (13), physical activity level using the Short Form International Physical Activity Questionnaire (IPAQ-sf), and Forced Expiratory Volume over 1 second (FEV$_1$).
During the three CPET sessions breath-by-breath data were collected for peak expired ventilator volumes (VO2peak), maximum work rate (Watt), total test time, peak carbon dioxide output (CO2), and Respiratory Exchange Ratio (RER) (CS-200, Schiller AG, Baar, Switzerland). Furthermore, maximum heart rate (beats per minute) was collected using a pulse watch (Polar HR RS 400, Polar Electro, Kempele, Finland). Immediately after the test, participants were asked to rate their maximum level of perceived exertion using the Borg 15-point scale (13).

**CPET Procedure**

The CPET was conducted on computer resisted arm cycle ergometers. Humidity, temperature, gas and volume calibration was performed before each test. Participants performed the CPET seated without their prostheses. During the test, participants were fastened around the hip, and floor contact was allowed for the non-amputated leg. The arm cranking center of rotation was in front of the sternum with an estimated 5-10 degrees of elbow flexion. Participants were given the opportunity to familiarize themselves with the equipment, without load, before starting the test. Each test started with a two minute warm-up period at 20 Watts resistance. The test started immediately after the warm up period at 40 Watts, and increased based on a ramp protocol at 20 Watts per minute. The participants were instructed to keep an asynchronous cranking cadence of 40-90 repetitions per minute (RPM), and the test stopped when the participant was volitional exhausted or was unable to constantly maintain a cranking cadence of 40 RPM. The protocol was tested on healthy able-bodied individuals, and based on these experiences we estimated a test time of 5-9 minutes (14, 15) and a peak respiratory exchange ratio (RER) of at least 1.10 or a self-reported rating of exertion ≥18 to be the threshold for a valid max-test (16). All data collection was obtained by the same trained tester (C.L.), who during all tests gave standardized verbal encouragement to the participants. The participants were blinded to all test parameters besides
cranking cadence and did not verbally or visually receive any indications of their test performance or results before, during or after completion of the tests.

**Data management and statistical analysis**

Two authors individually assessed all test results and subsequently cross-checked the data sheets to ensure data quality.

Continuous data were analysed for normal distribution using quantile-quantile-plots and if applicable, presented as means ± standard deviation (SD), and categorical data were presented as median and interquartile range (IQR). Due to a mask leakage, one VO2peak value from test 1 was estimated using multiple imputation by chained equations based on VO2peak values and maximal work rate (Watts) from all three tests (17). A sensitivity analysis was conducted comparing the ICC models with and without the imputed VO2peak value.

A two-way ANOVA with random effects of single measurements was used to calculate the Intraclass Correlation Coefficient (ICC). ICC\textsubscript{agreement} was calculated based on three test results which included residual variance plus variance due to systematic differences along with a 95% Confidence Interval (95%CI) of the ICC (18). ICC was interpreted as ≥0.40=poor agreement, 0.41-0.69=good agreement and ≤0.70=excellent agreement (19). The recommended value of 0.70 was used as the minimum standard for acceptable reliability (19).

To assess the absolute agreement for individual subjects, the Standard Error of Measurement (SEM\textsubscript{agreement}) was calculated based on the formula $SEM = sd * \sqrt{(1 - ICC)}$ (6).

The mean value of systematic bias and the 95% Limits of Agreement (LoA) were calculated using the formula mean ± 1.96 x SD\textsubscript{difference} and visually presented by Bland-Altman plots (20).

Twenty-five degrees of freedom has been suggested as sufficient for using ICC (21), corresponding to a sample size of 8 participants being tested three times. All statistical analyses were performed in
STATA 13.1 (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.) and level of statistical significance was set at 0.05 alpha level.
RESULTS

Eight male participants with unilateral traumatic lower limb amputations with a mean age of 32.5 (±4.57) years, who were amputated 6.25 (±2.82) years ago, without comorbidities, volunteered to participate in this study. Characteristics of the participants are presented in table I.

[Insert table I about here]

Results of the three CPET sessions are presented in table II. All tests sessions met the thresholds for a valid max-test with a test time between 5-9 minutes and a peak respiratory exchange ratio (RER) of at least 1.10 or a self-reported rating of exertion ≥18.

Test-retest reliability, standard error of the measurements and limits of agreement for VO2peak, total test time, maximum work rate, and maximum heart beat are presented in table 2. No statistically significant interaction effect was found between time and any of the outcome variables.

[Insert table II about here]

VO2peak

VO2peak showed a good but non-acceptable agreement with an ICC value of 0.51 with a standard error of the measurement of 0.18. A sensitivity-analysis showed no difference in interpretation of the ICC between the model without imputation ICC=0.50 and the model with imputation ICC=0.51.

Test time and maximum work rate

Total test time (seconds) and maximum work rate (watts) showed an excellent and acceptable agreement corresponding to ICC values of 0.74 and 0.73 respectively. Standard error of the measurement was for total test time 25.22 seconds and for maximum work rate 8.16 Watts.
Maximal heart rate

Test-retest reliability of maximum heart rate was found to be good but non-acceptable corresponding to an ICC of 0.64. The standard error of the measurement was found to be 6.19 beats per minute.

Limits of agreement

Figure 1a-d shows the Bland-Altman plots with 95% limits of agreement for VO2peak, maximum work rate, test time, and maximum heart rate for test 2 and 3 respectively. The mean difference showed a tendency to higher mean values at test three, with VO2peak being 0.13L higher, test time being 13.25 seconds longer, maximum work rate being 4.88 Watts higher and maximum heart beat being 0.88 beats per minute higher compared to test two.

Familiarization

We conducted a sensitivity analysis to investigate if a familiarization effect was present from the first to the latter two test sessions. This did not change the interpretation of the results for VO2peak, test time and maximum work rate. However, for maximal heart rate the interpretation changed from a good but non-acceptable ICC to an excellent and acceptable ICC (Supplementary Table S1).
DISCUSSION

In summary, the present study showed that an upper-body CPET designed specifically for testing aerobic capacity of younger adults with lower limb amputation showed excellent and acceptable test-retest reliability (ICC >0.70) for test time and maximum work rate. VO2peak (ICC=0.51) and heart rate (ICC=0.64) for all three test-sessions showed good but non-acceptable test-retest reliability.

**VO2peak**

VO2peak obtained during arm cycling only relates to 70% of the uptake obtained from testing on a stationary bicycle including lower limbs, hence the VO2peak was never expected to be similar to the VO2max of the participants (2). However, upper-body CPET is thought to be an alternative testing method for people with lower limb disabilities, showing acceptable test-retest reliability for VO2peak in existing literature in different patient populations (10, 15). On the contrary, we found a good but non-acceptable ICC for VO2peak, which might partly be explained by differences in the test protocol compared to former studies. In the present study, three test sessions were conducted, and the participants conducted a maximal test based on a fast ramp protocol designed specifically for younger lower limb amputees. It was estimated that the participants would reach a total test time of 5-9 minutes. This was successfully achieved, as the mean test time for all three test sessions was within this timeframe. The 5-9 minutes total test duration is shorter than the 8-12 minutes usually recommended for assessment of cardiopulmonary exercise testing performed with lower limbs (22). However, given the smaller amount of muscle mass involved in testing upper-body and the increased risk of peripheral fatigue, a shorter test time has previously been suggested when using arm crank testing in healthy younger adult participants (23) and in different patients groups (14, 15).
Test time, work rate and heart rate

The excellent and acceptable test-retest reliability for test time and work rate (Watts) found in the present study agrees with the literature regarding healthy subjects (10, 24).

Similarly, a mean maximum work rate of 155 Watts and maximal heart rate of 174 beats per minute are in line with formerly reported maximal values obtained during arm cycling in competitive male amputees of 147 Watts and 174 beats per minutes (25). Given that work rate and test duration are the main variables collected in routine practice, the excellent and acceptable reliability of this test protocol makes it applicable in numerous everyday clinical rehabilitation settings as no costly equipment for measuring gas exchange is needed. Simply having access to an arm cycle that can measure work rate is enough to be able to conduct a reliable maximum test.

SEM and LoA

The absolute reliability was investigated by calculating the SEM. SEM is a measure of how much test scores are spread around the “true” score (6). Since the formula for SEM contains the ICC these measures are closely related. In this specific group of participants, we found SEM values based on the latter two tests of 0.18 L/min for VO2peak, 8.16 watt for work rate, 25.22 seconds for total test time, and 6.19 beats/min for heart rate. Our results are therefore in line with formerly reported typical errors of 0.2 L/min for VO2peak, 5.8 Watts, and 4.2 beats/min obtained from maximal arm ergometer test in healthy adults (24). If the test is used in clinical practice or as an intervention effect parameter over time, clinicians should carefully consider the SEM as the errors nearly correspond to meaningful changes of 0.25L/min (1 METS/kilogram bodyweight) and 5 beats per minute.
The Bland-Altman plot is a simple way to evaluate the absolute agreement between two tests. Between the latter two tests we found a systematic bias of 0.13L for VO2peak with a limit of agreement of 0.53 L. The absolute agreement of maximal VO2 uptake is in general somewhat limited by high inter-individual variability (26), which is also present in the present study illustrated by the Bland-Altman plots. With a standardized test procedure including standardized feedback from the tester, we expected this variability to be lower. However, as the self-reported maximum subjective perceived exertion and RER values were above the threshold for a maximal test we did not have any reason not to include these findings and the variability must therefore be interpreted as inter-individual variability.

**Test protocol**

The physiological responses are found to be similar whether applying a step or a ramp protocol for arm crank ergometry (27). However, the evidence is inconclusive on whether an asynchronous (used in our study) or synchronous handlebar position gives higher maximal work rate, while peak oxygen uptake seems not to be affected by the choice of protocol. One study, that used a protocol starting a 0 Watts and increased by 10 Watts every minute, showed no difference in power output or VO2peak between asynchronous and synchronous arm cranking (28). On the other hand, another study used a protocol starting at 20 Watts and increasing 5 Watts every minutes found that the participants obtained higher lactate levels and prolonged exercise time when using the asynchronous protocol compared to the synchronous, but no statistical differences were observed for maximal heart rate or peak oxygen uptake (29). As the present study aimed to investigate reliability on a single protocol our results are therefore not influenced by differences in protocols. Nevertheless, it is unknown how differences in test protocols affect the test-retest reliability and our results might be limited to this specific test protocol.
Learning effect

Our results indicated no change in the interpretation of the results for VO2peak, test time and maximum work rate. However, for maximal heart rate the interpretation changed from a good but non-acceptable ICC to an excellent and acceptable ICC when the first test sessions were considered as a familiarization test. This could indicate that a learning effect could be present, but these results should be interpreted carefully considering the limited sample size.

The finding of a possible learning effect is contrary to what is seen when participants are tested on a stationary leg bike where a learning effect is not found (30). Usually only two tests are conducted in the existing literature on test-retest reliability of upper-body cardiopulmonary exercise testing, which makes it difficult to assess if a learning effect is present. To our knowledge this is the first study that has investigated the test-retest reliability using three test sessions for patients with lower limb disabilities. Hence, based on our findings the authors strongly recommend that patients conduct a formalized familiarization test before the actual test in both clinical and research settings.

Study strengths and limitations

The model used for calculating ICC was the ICC agreement for single measures, which is a strict model used for assessment of the absolute agreement including both variance due to systematic differences and variance due to error variance (31, 32). The strict ICC agreement model was used for analysis as this is how the test will be conducted and therefore interpreted on a daily basis.

A limitation of the current study is that we only investigated the intra-rater reliability by using the same data trained tester, who provided standardized verbal encouragements to the participants, and not the inter-rater reliability. It is unknown to what degree another trained tester providing standardized verbal feedback could influence the results of the test.

Although our main analysis met the required sample size for reliability, the learning effect sub-
analysis might be based on an insufficient number of observations. Furthermore, as we only included males with a mean age of 32.5 (4.57) years in this study our results fall short of being representative of the female population and other age groups.

The participants performed three tests at the same time of day, however, we did not control their dietary intake. Dietary intake influences the energy expenditure (33) but neither alterations in dietary carbohydrate or caffeine intake seems to influence the oxygen uptake or total test time when the test time is relatively short as in the present study (34, 35). A decision to not control dietary intake was made based on maintaining generalizability to everyday clinical practice.

In conclusion, we found good but non-acceptable test-retest reliability for VO2peak and maximal heart rate. However, we found acceptable and good test-retest reliability for work rate and total test time obtained from an arm cycle test specifically developed for younger adults with lower limb amputations. Furthermore, our findings indicate that familiarization with the test is important for reaching acceptable reliability.

DECLARATION OF CONFLICTING INTERESTS

The authors report no declarations of interest. The first author is funded by a strategic grant from the [Name of removed for blinded peer-review process].
REFERENCES


Table I: Characteristics of included participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean (±SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants with unilateral crus amputation (n=8)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>32.5 (4.57)</td>
</tr>
<tr>
<td>Years since amputation</td>
<td>6.25 (2.82)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>176.5 (6.00)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>73.25 (8.63)</td>
</tr>
<tr>
<td>Resting BP (systolic)</td>
<td>128.25 (8.55)</td>
</tr>
<tr>
<td>Resting BP (diastolic)</td>
<td>78.13 (8.04)</td>
</tr>
<tr>
<td>Resting HR (beats pr/min)</td>
<td>60.37 (4.00)</td>
</tr>
<tr>
<td>FEV₁ (liter)</td>
<td>3.76 (0.49)</td>
</tr>
<tr>
<td>FEV₁%</td>
<td>94.14 (7.20)</td>
</tr>
<tr>
<td>Physical activity level (Met-min/week)</td>
<td>3034.29 (1525.17)</td>
</tr>
</tbody>
</table>

Caption: Table I: Characteristics of included participants. BP=Blood Pressure, HR=Heat rate, FEV₁=Forced Expiratory Volume over 1 second, FEV₁%= Forced Expiratory Volume over 1 second in percentage of age adjusted norm values, Physical Activity measured by International Physical Activity Questionnaire (IPAQ), Met=metabolic equivalents
Table II: Arm crank test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Test 1 mean (±SD)</th>
<th>Test 2 mean (±SD)</th>
<th>Test 3 mean (±SD)</th>
<th>CV%</th>
<th>ICC for all three test (95%CI)</th>
<th>SEM</th>
<th>LoA test 2&amp;3 Mean (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VO2peak (L/min)</td>
<td>2.27 (0.26)</td>
<td>2.04 (0.40)</td>
<td>2.16 (0.23)</td>
<td>14.48</td>
<td>0.51 [0.11-0.85]</td>
<td>0.18</td>
<td>-0.13 ±(0.53)</td>
</tr>
<tr>
<td>Test time (min.sec)</td>
<td>5.43 (0.49)</td>
<td>5.45 (0.39)</td>
<td>5.58 (0.40)</td>
<td>7.83</td>
<td>0.74 [0.40-0.93]</td>
<td>0.25</td>
<td>-0.13 ±(0.54)</td>
</tr>
<tr>
<td>Maximum work rate (Watt)</td>
<td>153.13 (15.71)</td>
<td>154.25 (12.46)</td>
<td>159.13 (13.32)</td>
<td>8.95</td>
<td>0.73 [0.40-0.93]</td>
<td>8.16</td>
<td>-4.88 ±(17.25)</td>
</tr>
<tr>
<td>Maximum heart rate (beats per/min)</td>
<td>174.75 (10.31)</td>
<td>174.13 (6.97)</td>
<td>173.25 (12.16)</td>
<td>5.78</td>
<td>0.64 [0.21-0.90]</td>
<td>6.19</td>
<td>-0.88 ±(14.64)</td>
</tr>
<tr>
<td>Perceived exertion (Borg15)</td>
<td>17 (17-17)</td>
<td>17 (17-18)</td>
<td>18 (17-19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RER (ratio)</td>
<td>1.25 (1.15-1.30)</td>
<td>1.27 (1.18-1.34)</td>
<td>1.21 (1.15-1.37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Caption: Table II: Arm crank test results. Peak expired ventilator volumes (VO2peak), maximum subjective perceived exertion (Borg15), Respiratory Exchange Ratio (RER), Coefficient of Variation in percentage (CV%), ¤=Presented as Median and Inter-Quartile Range. Data is presented with one imputed observation in VO2peak (L) test 1.