Spinal Cord (2019) 57:669–678
https://doi.org/10.1038/s41393-019-0262-8

ARTICLE

Interrater and intrarater reliability of ventilatory thresholds determined in individuals with spinal cord injury

Ingrid Kouwijzer[1,2,3] · Rachel E. Cowan[4] · Jennifer L. Maher[4] · Floor P. Groot[5,6] · Feikje Riedstra[5,6] · Linda J. M. Valent[1] · Lucas H. V. van der Woude[2,7] · Sonja de Groot[2,3]

Received: 16 October 2018 / Revised: 11 February 2019 / Accepted: 11 February 2019 / Published online: 28 February 2019

Abstract

Study design Cross-sectional.

Objectives Individualized training regimes are often based on ventilatory thresholds (VTs). The objectives were to study: (1) whether VTs during arm ergometry could be determined in individuals with spinal cord injury (SCI), (2) the intrarater and interrater reliability of VT determination.

Setting University research laboratory.

Methods Thirty graded arm crank ergometry exercise tests with 1-min increments of recreationally active individuals (tetraplegia ($N = 11$), paraplegia ($N = 19$)) were assessed. Two sports physicians assessed all tests blinded, randomly, in two sessions, for VT1 and VT2, resulting in 240 possible VTs. Power output (PO), heart rate (HR), and oxygen uptake (VO_2) at each VT were compared between sessions or raters using paired samples *t*-tests, Wilcoxon signed-rank tests, intraclass correlation coefficients (ICC, relative agreement), and Bland–Altman plots (random error, absolute agreement).

Results Of the 240 VTs, 217 (90%) could be determined. Of the 23 undetermined VTs, 2 (9%) were VT1 and 21 (91%) were VT2; 7 (30%) among individuals with paraplegia, and 16 (70%) among individuals with tetraplegia. For the successfully determined VTs, there were no systematic differences between sessions or raters. Intrarater and interrater ICCs for PO, HR, and VO_2 at each VT were high to very high (0.82–1.00). Random error was small to large within raters, and large between raters.

Conclusions For VTs that could be determined, relative agreement was high to very high, absolute agreement varied. For some individuals, often with tetraplegia, VT determination was not possible, thus other methods should be considered to prescribe exercise intensity.

Introduction

In wheelchair users with spinal cord injury (SCI) cardiorespiratory fitness is generally reduced [1]. Low cardiorespiratory fitness and low levels of physical activity are shown to be associated with high prevalence of cardiometabolic disease, which is the leading cause of mortality in this population [2, 3]. To increase cardiorespiratory fitness, exercise interventions such as handcycling may be introduced during or after rehabilitation [4–6]. To promote handcycling in the Netherlands and to increase cardiorespiratory fitness after rehabilitation, an annual handcycle race called the HandbikeBattle [7, 8] has been held since 2013 in Austria. To optimally train for events like this, but also in or after rehabilitation in general, individualized training schemes are required.

✉ Ingrid Kouwijzer
i.kouwijzer@heliomare.nl

[1] Research and Development, Heliomare Rehabilitation Center, Wijk aan Zee, the Netherlands

[2] University of Groningen, University Medical Center Groningen, Center for Human Movement Sciences, Groningen, the Netherlands

[3] Amsterdam Rehabilitation Research Center | Reade, Amsterdam, the Netherlands

[4] Department of Neurological Surgery, Miller School of Medicine & The Miami Project to Cure Paralysis, University of Miami, Miami, FL, USA

[5] Heliomare Rehabilitation Center, Wijk aan Zee, the Netherlands

[6] Sport- en Beweegkliniek, Haarlem, the Netherlands

[7] University of Groningen, University Medical Center Groningen, Center for Rehabilitation, Groningen, the Netherlands

Individualized training schemes can be based on results of a graded exercise test (GXT). Training prescriptions based on maximum values, such as percentage maximum heart rate (HR) or power output (PO), are common, as well as prescriptions based on percentage HR reserve or ventilatory thresholds (VTs) [9]. Training intensity prescription based on relative percentages is shown to have downsides in able-bodied individuals. It seems not to take into account the individual's metabolic response to exercise, and has shown less improvements in maximum oxygen uptake (VO_2 max) after training compared with training intensity based on VTs [9, 10]. Therefore, prescribing training intensities based on VTs may more reliably achieve fitness gains. The first ventilatory threshold (VT1) is a physiological point during exercise at which a nonlinear increase in carbon dioxide (CO_2) production occurs, coinciding with the first increase in lactate production [11]. The second ventilatory threshold (VT2) represents the onset of exercise-induced hyperventilation with respect to VCO_2 as a reaction to metabolic acidosis, which coincides with the maximal lactate steady state [11, 12]. These VTs provide boundaries that allow to set individualized training zones: zone 1 at low intensity (below VT1), zone 2 at moderate intensity (between VT1 and VT2), and zone 3 at high intensity (above VT2) [12, 13]. This training principle has been developed in studies on lower-body exercise with able-bodied participants and athletes, and little or no research has been done regarding the reliability of VT determination in upper-body GXT in individuals with SCI. Therefore, the question arises whether the reliability of determination of both VTs is sufficient to set training schemes for individuals with SCI.

For able-bodied leg exercise, VT1 is normally positioned at 50–60% peak oxygen uptake (VO_2 peak) and VT2 at 70–80% VO_2 peak [14]. This is, however, dependent on cardiorespiratory fitness as values for VT1 and VT2 could increase to 75 and 90% VO_2 peak for elite endurance athletes [12]. Studies in able-bodied cycling showed that experienced raters are able to identify VT1 in 90–94% of participants [15, 16]. Intrarater reliability of VT1 determination was high (intraclass correlation coefficient (ICC) 0.97) in one study [17], whereas interrater reliability varied (ICC 0.21–0.98) within and between studies [16, 17]. The identification rate and reliability of VT2 identification are largely unknown; only one study reports on this topic with an intrarater reliability (ICC) of 0.94–0.96 and an interrater reliability of 0.81–0.91 [18].

However, few studies reported on VTs during upper-body exercise in individuals with SCI. In two studies, 89–96% of VT1 and 74% of VT2 could be determined in wheelchair athletes with SCI [19, 20]. In both studies almost all undetermined VTs appeared to involve athletes with tetraplegia. Leicht et al. [19] explained that for athletes with tetraplegia the percentage of identifiable VT2s might be lower compared with able-bodied athletes, as the absolute ventilatory responses are generally low, resulting in a narrower range of ventilatory values compared with able-bodied athletes. A very recent study supports these findings as VT1 was only identified in 68% of untrained individuals with tetraplegia [21]. For the VT1, Coutts et al. reported a (Pearson) correlation of 0.95 between two raters for athletes with paraplegia and tetraplegia [20], and Bhambani et al. reported a Pearson correlation of 0.90 between two raters for trained and untrained individuals with tetraplegia [22]. However, although ICCs are more appropriate to assess intrarater and interrater reliability than Pearson correlations, they were not reported.

Unfortunately, no studies reported on reliability of VT determination for both thresholds, investigated in a non-athlete population with SCI. Therefore, it remains unclear whether VTs can be used to set individualized training schemes in this less fit population. The aims were, therefore the following:

1. To examine whether it is possible to detect both VTs in recreationally active individuals with tetraplegia or paraplegia.
2. To examine the interrater and intrarater reliability of VT determination.

Methods

The present study was a retrospective study: the data of the GXTs with 1-min increments of a previous study by Maher et al. [23] were re-analyzed to answer the research questions. Two sports physicians independently evaluated the tests twice during two separate sessions.

Participants

Thirty-three recreationally active individuals with SCI were recruited to participate in the study: 19 individuals with paraplegia and 14 with tetraplegia, 28 men, age: 38 ± 10 years, time since injury (TSI): 12 ± 9 years, body mass: 76 ± 19 kg, height: 1.75 ± 0.08 m. They were recruited through the Miami Project to Cure Paralysis database and voluntarily trained at the Miami Project gym at least once a week. Inclusion criteria: age ≥ 18 years, non-progressive SCI, TSI of at least 6 months, and self-reported inability to use lower extremity contractions to assist in transfers. Exclusion criteria: angina or myocardial infarction within the last month or pain in the upper extremities [23]. Informed consent was obtained from all participants included in the study. The study was approved by the

University of Miami Institutional Review Board, Miami, United States of America.

Test procedure

All GXTs were performed with an (asynchronous) arm crank ergometer (Lode Angio, Groningen, the Netherlands). Participants performed the tests in their own wheelchair; positioned with arms slightly flexed in the furthest horizontal position; participants with tetraplegia used hand wraps to ensure a tight grip on the cranks; and wedges were used to minimize the movement of the wheelchair. As individualized protocols are preferred for individuals with SCI [24, 25], the starting workload and step size of every participant were individualized based on questions regarding activity level, current fitness program, and the ability to perform a floor-to-chair transfer [23]. The aim was to develop an individualized 1-min stepwise protocol with a duration between 8 and 12 min [26]. This resulted in an individualized starting workload of 5–90 W and step size of 10 W for participants with paraplegia, and start workload of 5–30 W and step size of 3–10 W for participants with tetraplegia. The prescribed cadence was between 60 and 65 rpm. Criteria to stop the test were volitional exhaustion or failure in keeping a constant cadence above 55 RPM. During the test, PO (W) was continuously measured. Gas exchange was measured breath-by-breath (Vmax Encore metabolic cart, Carefusion, Vyaire Medical, Mettawa, IL, USA) and HR was measured by standard 12-lead electrocardiography. The metabolic cart was calibrated before each test. All raw data, except for PO, were processed using a moving average over a 15-breath window [27]. VO_2 peak and HR peak were defined as the highest 15-breath average of VO_2 and HR, respectively. PO peak and the PO at each VT were defined as the last completed work rate step, plus half times the work rate increment for any 30-s block in the non-completed work rate step [28].

Determination of ventilatory thresholds

All data of the GXTs were represented in plots as described by Wasserman et al. [29] via a custom-made Matlab script according to the preferences of both raters [Matlab R2012b, Mathworks Inc., Natick, MA, USA]. Three plots were presented to the raters: (1) VCO_2 versus VO_2, (2) the ventilatory equivalents of oxygen (Ve/VO_2) and carbon dioxide (Ve/VCO_2) versus time, and (3) respiratory exchange ratio (RER) versus time. VT1 was defined as an increase in slope of more than one in the first plot (V-slope method) [12, 15, 19, 30], and as the first sustained rise in Ve/VO_2 without a concomitant increase in Ve/VCO_2, in the second plot (ventilatory equivalents method) [15, 19, 30, 31]. The RER plot was used as extra reference [12, 30, 31]. VT2 was defined as the first sustained increase in Ve/VCO_2 (ventilatory equivalents method), in the plot with Ve/VO_2 and Ve/VCO_2 versus time [12, 14, 31], and as second increase in slope in the plot with VCO_2 versus VO_2 [12, 18]. Again, the RER plot was used as extra reference; for example for the raters to be certain that RER at VT2 was higher than RER at VT1 [12, 30, 31]. The raters assessed all three plots for each VT and made their final decision based on the V-slope or the ventilatory equivalents, depending on which plot most clearly showed that particular VT.

Two experienced sports physicians independently and randomly assessed the sets of graphs. They had at least 4 years of experience with VT determination in able-bodied athletes and in upper-body exercise in individuals with a disability. They were blinded to participant ID and injury level. For each determined VT, the Matlab script calculated the corresponding PO, HR, VO_2, and RER at that threshold. When a rater thought that a VT was indeterminate, the test data for that VT were rejected. To calculate intrarater reliability, both raters assessed all tests twice (in different random order) with at least 1 week in between.

Statistical analysis

Statistical analyses were performed using SPSS (IBM SPSS Statistics 20, SPSS, Inc, Chicago, IL, USA). The data were tested for normality using Kolmogorov–Smirnov tests with Lilliefors Significance Correction and Shapiro–Wilk tests. Additionally, z-scores for skewness and kurtosis were calculated. To assess intrarater reliability for each VT, PO, HR, and VO_2 at that VT were compared between the first and second session. To assess interrater reliability for each VT, PO, HR, and VO_2 at that VT were compared between rater one and two for the first session. Systematic differences were investigated with paired samples t-tests for the total group and Wilcoxon signed-rank tests and Mann–Whitney tests within subgroups (tetraplegia and paraplegia) as data within subgroups were not normally distributed. ICCs with 95% confidence intervals (CI) were used to measure relative agreement on group level (ICC, two-way random, absolute agreement, single measures). For clinical/training purposes, Bland–Altman plots with 95% limits of agreement (LoA) were used to measure absolute agreement on an individual level [32]. The following interpretation was used for the ICC: 0.00–0.25, little to no correlation; 0.26–0.49, low correlation; 0.50–0.69, moderate correlation; 0.70–0.89, high correlation; and 0.90–1.00, very high correlation [33]. Values were considered significant at $p < 0.05$.

Results

Due to technical problems and short periods of stopping during testing, a total of three tests were excluded. This resulted in 30 tests to be assessed (tetraplegia $N = 11$, paraplegia $N = 19$). These 30 tests, with two possible VTs each, were assessed during two sessions by two independent raters, resulting in a total of 240 VTs to be analyzed (30 tests × 2 VTs × 2 sessions × 2 raters). For two tests, HR data were excluded due to problems with the HR monitoring system. The test peak values are shown in Table 1.

Determination of ventilatory thresholds

Of the 240 VTs to be analyzed, 217 VTs (90%) could be determined. Of the 23 undetermined VTs, 2 (9%) were VT1 and 21 (91%) were VT2; and 7 (30%) related to tests in individuals with paraplegia and 16 (70%) to tests in individuals with tetraplegia (Fig. 1). In 18 out of the 30 tests (60%), both VTs could be determined during both sessions by both raters. Fourteen of these tests were related to individuals with paraplegia. Among individuals with paraplegia, there were no differences in peak test physiological values between tests where all VTs could ($N = 14$) and could not ($N = 5$) be determined (*Median* (*Mdn*) ± *standard error* (*SE*): VO_2 peak 1.50 ± 0.17 L/min vs 1.11 ± 0.25 L/min, $p = 0.19$; PO peak 98 ± 10 W vs 70 ± 15 W, $p = 0.11$; HR peak 161 ± 6.8 bpm vs 156 ± 5 bpm, $p = 0.20$; RER peak 1.29 ± 0.02 vs 1.43 ± 0.08, $p = 0.39$). However, test duration was significantly lower in tests where one or more VTs could not be determined (*Mdn* 5.1 ± 0.6 min), compared with tests where all VTs could be determined (*Mdn* 7.6 ± 0.5 min, $U = 11$, $z = -2.22$, $p = 0.026$). Four out of five individuals, of whom one or both VTs could not be determined by one or both raters, were individuals with a high paraplegia (thoracic level 1–5).

Table 1 Arm crank test peak values ($N = 30$)

	Total group		Paraplegia		Tetraplegia	
	N	M ± SD	N	M ± SD	N	M ± SD
PO peak (W)	30	73 ± 41	19	92 ± 38	11	40 ± 20
VO_2 peak (L/min)	30	1.23 ± 0.65	19	1.50 ± 0.64	11	0.76 ± 0.32
RER peak	30	1.28 ± 0.12	19	1.30 ± 0.12	11	1.25 ± 0.12
HR peak (bpm)	28	140 ± 30	17	158 ± 21	11	112 ± 17
Test duration (min)	30	6.5 ± 2.2	19	7.1 ± 2.0	11	5.4 ± 2.1

PO peak = peak power output, VO_2 peak = peak oxygen uptake, RER peak = peak respiratory exchange ratio, HR peak = peak heart rate

Among individuals with tetraplegia, there were no differences in peak test physiological values and test duration between tests where all VTs could ($N = 4$) and could not ($N = 7$) be determined (VO_2 peak 0.79 ± 0.09 L/min vs 0.77 ± 0.15 L/min, $p = 0.79$; PO peak 44 ± 11 W vs 35 ± 8 W, $p = 0.65$; HR peak 118 ± 14 bpm vs 113 ± 3 bpm, $p = 0.79$; RER peak 1.30 ± 0.04 vs 1.22 ± 0.06, $p = 0.53$; test duration 5.6 ± 1.4 min vs 4.8 ± 0.6 min, $p = 0.53$).

Intrarater reliability

For the total group and injury subgroups no systematic differences were found between session 1 and 2, except for the VO_2 at VT2 for the group with paraplegia in rater 1 (Δ Median: 0.00 L/min, Δ Mean: 0.06 L/min, T 7.0, SE 12.7, $p = 0.01$). Tables 2–4 show the intrarater reliability for the total group and subgroups. The relative agreement between rating sessions was very high for both raters. In subgroups, the relative agreement varied between high to very high for both raters, although small sample size and unidentifiable VTs have reduced the statistical power. This can especially be seen in Table 4, where 95% CI were wide despite the high to very high ICC. Bland–Altman plots showed small systematic error as represented by small mean differences. Random error was small to large as represented by the small to wide 95% LoA in Figs. 2 and 3. Fig. 2 a, b, d, e and Fig. 3 a, b, d, e show the absolute agreement within raters for PO and HR, respectively.

Interrater reliability

There were no systematic differences between rater 1 and rater 2. The relative agreement between both raters was high to very high for the total group as well as for the subgroups (Tables 2, 3, 4). Again, due to small sample sizes and the number of excluded undetermined VTs, the number of tests in the subgroups was small. Bland–Altman plots showed

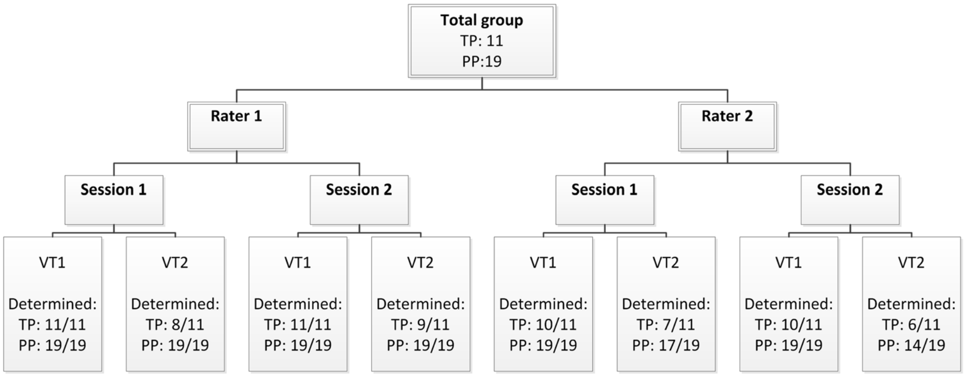

Fig. 1 Flowchart of the thresholds that could be determined by two experienced raters in 30 individuals with spinal cord injury during arm crank ergometry. TP = group with tetraplegia ($N = 11$), PP = group with paraplegia ($N = 19$), VT = ventilatory threshold

Table 2 Threshold characteristics rater 1 and rater 2 for the total group of participants during arm crank testing

Total group (N = 30)														
Rater 1							Rater 2							
	Session 1		Session 2		Intra		Session 1		Session 2		Intra		Inter	
	N	M ± SD	N	M ± SD	N	ICC (95%CI)	N	M ± SD	N	M ± SD	N	ICC (95% CI)	N	ICC (95% CI)
PO at VT1 (W)	30	29 ± 20	30	30 ± 21	30	0.94 (0.88–0.97)	29	31 ± 21	29	29 ± 21	29	0.98 (0.96–0.99)	29	0.96 (0.92–0.98)
% of PO peak		38 ± 15		39 ± 16				38 ± 16		36 ± 15				
PO at VT2 (W)	27	50 ± 32	28	49 ± 29	27	0.99 (0.98–1.00)	24	58 ± 28	20	65 ± 29	20	0.94 (0.86–0.98)	22	0.96 (0.90–0.98)
% of PO peak		64 ± 15		63 ± 15				74 ± 11		78 ± 9				
VO_2 at VT1 (L/min)	30	0.65 ± 0.25	30	0.65 ± 0.25	30	0.96 (0.92–0.98)	29	0.68 ± 0.24	29	0.66 ± 0.23	29	0.98 (0.95–0.99)	29	0.95 (0.90–0.98)
% of VO_2 peak		59 ± 18		59 ± 17				59 ± 17		58 ± 15				
VO_2 at VT2 (L/min)	27	0.92 ± 0.50	28	0.88 ± 0.43	27	0.97 (0.93–0.99)	24	1.03 ± 0.43	20	1.11 ± 0.43	20	0.96 (0.91–0.99)	22	0.97 (0.94–0.99)
% of VO_2 peak		74 ± 16		73 ± 16				82 ± 14		84 ± 13				
HR at VT1 (bpm)	28	105 ± 18	28	105 ± 18	28	0.94 (0.87–0.97)	27	104 ± 16	27	104 ± 17	27	1.00 (0.99–1.00)	27	0.95 (0.89–0.98)
% of HR peak		77 ± 11		77 ± 11				75 ± 10		75 ± 10				
HR at VT2 (bpm)	25	119 ± 22	26	118 ± 22	25	0.99 (0.97–1.00)	22	123 ± 24	18	127 ± 27	18	0.94 (0.84–0.98)	20	0.89 (0.74–0.95)
% of HR peak		84 ± 10		84 ± 10				88 ± 9		90 ± 9				

PO = power output, VO_2 = oxygen uptake, HR = heart rate, VT1 = first ventilatory threshold, VT2 = second ventilatory threshold. M = mean, SD = standard deviation, N = number of tests, ICC = intraclass correlation coefficient, 95% CI = 95% confidence intervals. Intra = the intrarater reliability, inter = the interrater reliability. The interrater reliability is based on session 1 of both raters. All ICCs were significant

Table 3 Threshold characteristics rater 1 and rater 2 for the group with paraplegia during arm crank testing

Paraplegia group (N = 19)														
Rater 1							Rater 2							
	Session 1		Session 2		Intra		Session 1		Session 2		Intra		Inter	
	N	M ± SD	N	M ± SD	N	ICC (95% CI)	N	M ± SD	N	M ± SD	N	ICC (95% CI)	N	ICC (95% CI)
PO at VT1 (W)	19	37 ± 20	19	40 ± 20	19	0.91 (0.79–0.97)	19	38 ± 21	19	36 ± 21	19	0.97 (0.93–0.99)	19	0.95 (0.89–0.98)
% of PO peak		38 ± 13		42 ± 14				39 ± 15		37 ± 15				
PO at VT2 (W)	19	61 ± 30	19	59 ± 27	19	0.98 (0.95–0.99)	17	68 ± 26	14	77 ± 26	14	0.91 (0.74–0.97)	17	0.95 (0.87–0.98)
% of PO peak		64 ± 12		63 ± 13				73 ± 11		77 ± 9				
VO_2 at VT1 (L/min)	19	0.75 ± 0.24	19	0.76 ± 0.24	19	0.94 (0.86–0.98)	19	0.75 ± 0.25	19	0.74 ± 0.23	19	0.97 (0.92–0.99)	19	0.95 (0.88–0.98)
% of VO_2 peak		53 ± 13		54 ± 14				53 ± 14		52 ± 13				
VO_2 at VT2 (L/min)	19	1.09 ± 0.48	19	1.04 ± 0.41	19	0.95 (0.88–0.98)*	17	1.18 ± 0.42	14	1.27 ± 0.41	14	0.94 (0.84–0.98)	17	0.97 (0.91–0.99)
% of VO_2 peak		73 ± 14		71 ± 16				80 ± 14		82 ± 13				
HR at VT1 (bpm)	17	111 ± 17	17	112 ± 18	17	0.92 (0.79–0.97)	17	110 ± 15	17	110 ± 16	17	1.00 (0.99–1.00)	17	0.95 (0.88–0.98)
% of HR peak		71 ± 8		72 ± 10				70 ± 7		70 ± 7				
HR at VT2 (bpm)	17	127 ± 20	17	126 ± 20	17	0.98 (0.95–0.99)	15	134 ± 20	12	139 ± 23	12	0.88 (0.66–0.97)	15	0.82 (0.55–0.94)
% of HR peak		81 ± 9		80 ± 10				85 ± 10		87 ± 10				

PO = power output, VO_2 = oxygen uptake, HR = heart rate, VT1 = first ventilatory threshold, VT2 = second ventilatory threshold. M = mean, SD = standard deviation, N = number of tests, ICC = intraclass correlation coefficient, 95%CI = 95% confidence intervals. Intra = the intrarater reliability, inter = the interrater reliability. The interrater reliability is based on session 1 of both raters. All ICCs were significant. * = outcome of the Wilcoxon signed-rank test for systematic differences, significant at $p < 0.05$

small systematic error as represented by small mean differences. Random error was generally large as represented by wide 95% LoA in Figs. 2 and 3. Fig. 2 c, f and Fig. 3 c, f show the absolute agreement between raters for PO and HR, respectively.

Discussion

Of all VTs to be analyzed, 90% could be determined. Of the undetermined VTs, most were VT2 and related to individuals with tetraplegia. In 60% of the tests, both thresholds could be determined during both sessions by both raters. For the successfully determined VTs, the relative intrarater reliability was very high, whereas random error ranged from small to large within raters and among outcome measures. The relative interrater reliability was high to very high with a low absolute agreement due to large random error.

The participants of the present study were recreationally active individuals with SCI. For physical fitness, the participants with paraplegia scored "good" for VO_2 peak

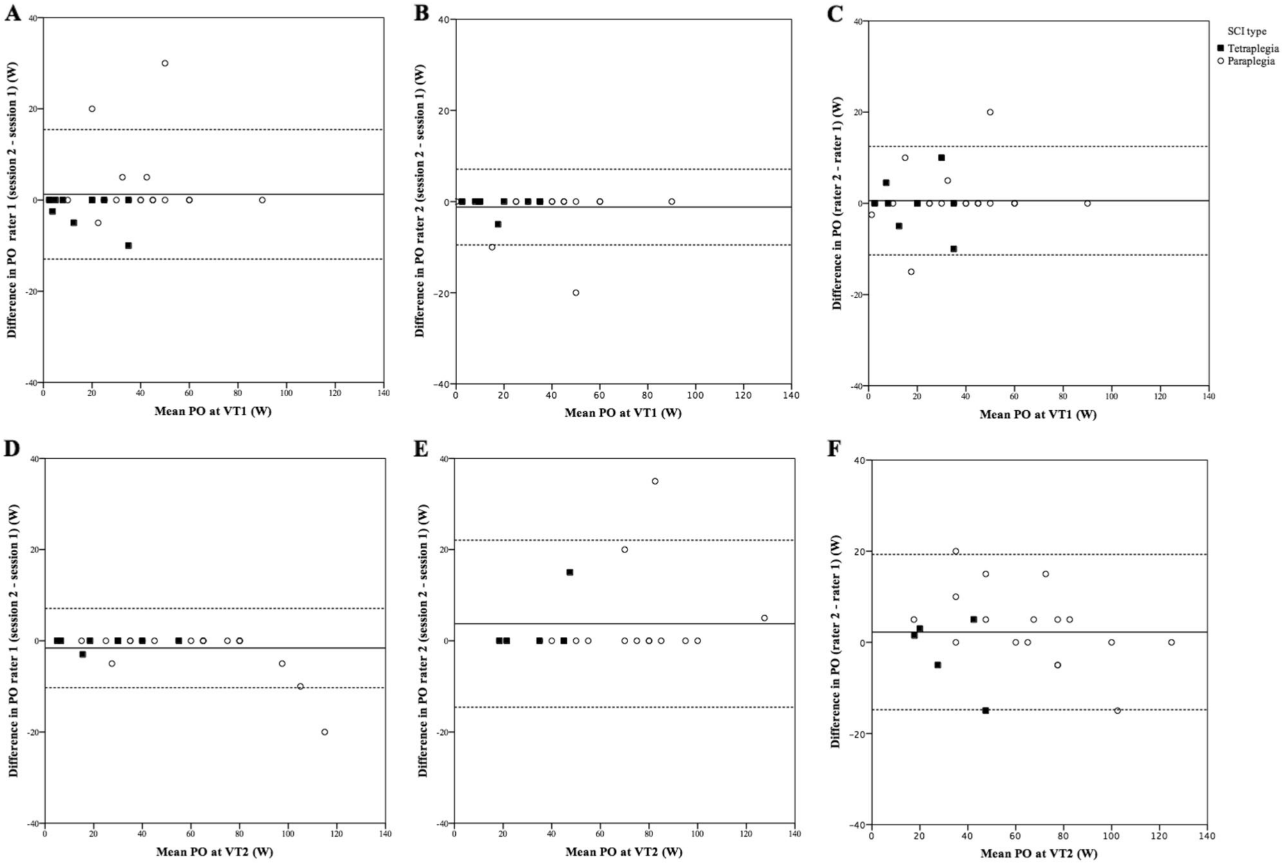

Fig. 2 Bland–Altman plot representing the absolute agreement of the power output (PO) within raters and between raters. Solid line represents the mean bias (systematic error), dotted lines represent mean ± 2 SD (95% LoA; random error). Circles and squares represent individuals with paraplegia and tetraplegia, respectively. **a** Intrarater reliability rater one at VT1. **b** Intrarater reliability rater two at VT1. **c** Interrater reliability at VT1. **d** Intrarater reliability rater one at VT2. **e** Intrarater reliability rater two at VT2. **f** Interrater reliability at VT2

compared with the average untrained population with paraplegia, based on the study by Simmons et al. [34]. The participants with tetraplegia scored “average” for VO_2 peak compared with the average untrained population with tetraplegia [34]. We might conclude that the population in the present study has a physical fitness somewhat above average, compared with the untrained population with SCI.

For the individuals with paraplegia, VO_2 at VT1 and VT2 was on average 53 and 76% of VO_2 peak, respectively. In previous literature across all test modes and fitness levels, VT1 has been reported as occurring at between 56 and 77% of VO_2 peak in individuals with paraplegia [19–21, 35, 36], whereas VT2 has been reported at 78% of VO_2 peak [19]. Possible small differences between the present study and previous literature might be explained by mode of exercise and training status of the participants. Physical fitness of the studied population in previous literature is generally higher than in the present study (VO_2 peak on average 1.9 L/min in previous literature vs 1.5 L/min in present study) [19, 35, 36].

For individuals with tetraplegia, VO_2 at VT1 and VT2 was on average 68 and 81% of VO_2 peak, respectively. In previous literature across all test modes and fitness levels, VT1 has been reported as occurring at between 63 and 87% of VO_2 peak in individuals with tetraplegia [19–22], whereas VT2 has been reported at 75% of VO_2 peak [19]. Overall, VTs of this subgroup are comparable with those reported in literature.

Ninety percent of VTs could be determined in the present study. This is comparable with literature with able-bodied participants [15, 16] and athletes with SCI [19, 20]. Most of the VTs that could not be determined were VT2s and related to tests in individuals with tetraplegia. Leicht et al. [19] found comparable results; two out of 19 VT1s (11%) could not be determined, both in athletes with tetraplegia, and 5 out of 19 VT2s (26%) could not be determined, of which 3 belonged to athletes with tetraplegia. Leicht et al. [19] explained their findings by lower absolute ventilatory responses in individuals with SCI, and tetraplegia specifically, resulting in a narrower range of ventilatory values compared with able-bodied athletes. In the present study there was no significant difference in VO_2 peak between individuals with tetraplegia whose VTs could be determined compared to those whose VTs could not be determined.

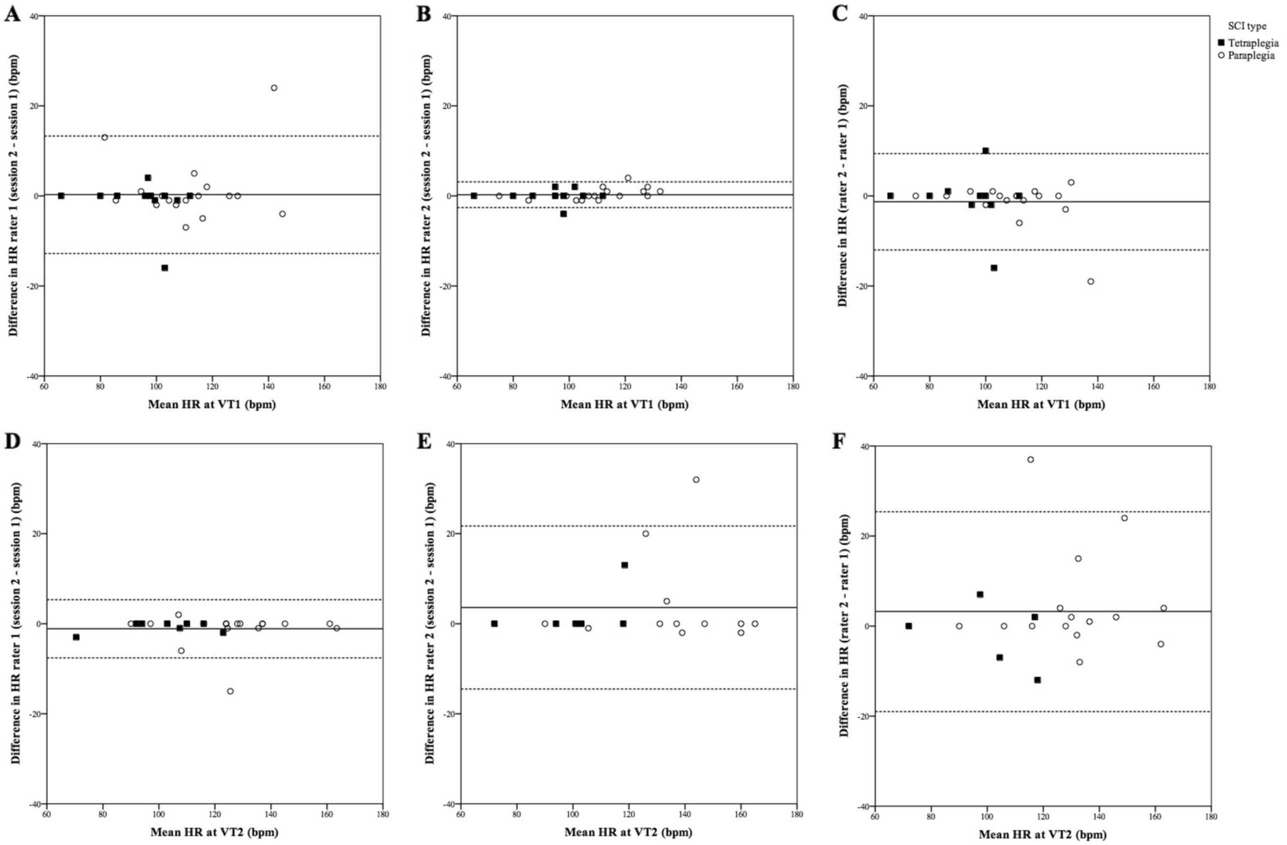

Fig. 3 Bland–Altman plot representing the absolute agreement of the heart rate (HR) within raters and between raters. Solid line represents the mean bias (systematic error), dotted lines represent mean ± 2 SD (95% LoA; random error). Circles and squares represent individuals with paraplegia and tetraplegia, respectively. **a** Intrarater reliability rater one at VT1. **b** Intrarater reliability rater two at VT1. **c** Interrater reliability at VT1. **d** Intrarater reliability rater one at VT2. **e** Intrarater reliability rater two at VT2. **f** Interrater reliability at VT2

However, although not significant, which is potentially due to small sample sizes, it can be seen that for both persons with tetraplegia and paraplegia PO peak and test duration were generally lower in tests where one or both VTs could not be determined. This also might explain the finding that the proportion of undetermined VTs was higher in individuals with tetraplegia compared with individuals with paraplegia. This is supported by a recent study, where in 32% of tests, VT1 could not be determined in individuals with tetraplegia [21]. They explain their findings by lower peak cardiorespiratory responses and lower test duration for those individuals, compared with tests where VT1 could be determined. Another reason for not being able to determine VTs in untrained individuals with SCI, especially at higher intensity (VT2), might be premature termination of the test due to peripheral fatigue. In the present study, three out of twelve individuals, where one or both VTs could not be determined, stated that fatigue in the arms was the reason to stop the test.

For the VTs that could be determined, relative agreement for the total group within and between raters was high to very high. The SCI subgroups results might be hard to interpret, as these groups were small. The results are comparable to previous literature with able-bodied participants and wheelchair athletes, where an intrarater reliability of 0.94–0.97 was reported [17, 18] and an interrater reliability of 0.81–0.95 [18, 20, 22]. The absolute agreement varied between outcome measures. For some measures, such as HR at VT2 between raters, the random error was large, as depicted in Fig. 3f. This figure also shows a certain degree of heteroscedasticity: random error appears to be larger for individuals with a higher HR at VT2, i.e., those with a paraplegia.

On group level the agreement is high to very high, but on individual level there might be large differences between rating sessions or raters, which has large implications for the correct prescription of exercise intensity of that individual. This suggests that relative agreement of VT determination should be interpreted with caution, not only in the present study, but also in previous literature, as the absolute agreement was unfortunately often not reported.

Practical applications

On group level the results of the present study are positive. For the majority of tests, the VTs could be determined and

Table 4 Threshold characteristics rater 1 and rater 2 for the group with tetraplegia during arm crank testing

Tetraplegia group ($N = 11$)

	Rater 1						Rater 2							
	Session 1		Session 2		Intra		Session 1		Session 2		Intra		Inter	
	N	M ± SD	*N*	M ± SD	*N*	ICC (95% CI)	*N*	M ± SD	*N*	M ± SD	*N*	ICC (95% CI)	*N*	ICC (95% CI)
PO at VT1 (W)	11	16 ± 13	11	15 ± 12	11	0.96 (0.86–0.99)	10	17 ± 13	10	17 ± 13	10	0.99 (0.97–1.00)	10	0.93 (0.73–0.98)
% of PO peak		37 ± 19		32 ± 18				36 ± 19		35 ± 17				
PO at VT2 (W)	8	25 ± 17	9	28 ± 18	8	1.00 (0.99–1.00)	7	33 ± 11	6	37 ± 14	6	0.89 (0.49–0.98)	5	0.85 (0.15–0.98)
% of PO peak		63 ± 20		64 ± 19				74 ± 11		79 ± 8				
VO_2 at VT1 (L/min)	11	0.48 ± 0.19	11	0.47 ± 0.16	11	0.97 (0.90–0.99)	10	0.53 ± 0.15	10	0.52 ± 0.16	10	0.99 (0.97–1.00)	10	0.89 (0.61–0.97)
% of VO_2 peak		69 ± 20		67 ± 19				69 ± 18		67 ± 16				
VO_2 at VT2 (L/min)	8	0.51 ± 0.21	9	0.55 ± 0.25	8	0.98 (0.89–1.00)	7	0.68 ± 0.18	6	0.73 ± 0.18	6	0.98 (0.88–1.00)	5	0.88 (0.21–0.99)
% of VO_2 peak		77 ± 21		76 ± 18				87 ± 12		89 ± 14				
HR at VT1 (bpm)	11	96 ± 14	11	95 ± 13	11	0.93 (0.78–0.98)	10	94 ± 13	10	94 ± 13	10	0.99 (0.97–1.00)	10	0.90 (0.66–0.98)
% of HR peak		86 ± 7		85 ± 8				84 ± 7		84 ± 7				
HR at VT2 (bpm)	8	102 ± 16	9	102 ± 16	8	1.00 (0.98–1.00)	7	100 ± 15	6	102 ± 19	6	0.95 (0.75–0.99)	5	0.93 (0.54–0.99)
% of HR peak		91 ± 8		91 ± 7				93 ± 5		95 ± 3				

PO = power output, VO_2 = oxygen uptake, HR = heart rate, VT1 = first ventilatory threshold, VT2 = second ventilatory threshold. M = mean, SD = standard deviation, N = number of tests, ICC = intraclass correlation coefficient, 95%CI = 95% confidence intervals. Intra = the intrarater reliability, inter = the interrater reliability. The interrater reliability is based on session 1 of both raters. All ICCs were significant

relative agreement within and between raters was high to very high. Nevertheless, for 7 out of 11 tests of individuals with tetraplegia, one or both raters could not determine one or both VTs. This seemed to coincide with short test duration. Despite the extensive experience of the testers with testing in individuals with SCI, it was difficult to select a protocol resulting in test duration between 8 and 12 min. It must be emphasized that individualized protocol selection is important for individuals with SCI. However, optimal protocol selection is comprehensive as cardiorespiratory fitness in individuals with SCI is based on a lot of factors, such as lesion level, sex, BMI, and training status [24, 25, 34]. As such, tests with a duration less than 8 min are common in clinical practice and are not specific for the present study [21].

As known, training intensity based on HR peak or HR reserve might not be applicable to individuals with a lesion level above thoracic spinal nerve 6 due to the altered sympathetic response to exercise [37], this is also shown in the present study, as HR peak was low in individuals with tetraplegia (Table 1). The present study shows that it is sometimes impossible to determine VTs in this group, which makes training based on training zones challenging as well. Other methods to prescribe exercise intensity might provide better precision, such as training based on ratings of perceived exertion and/or %PO peak [38]. In the present study, it was not investigated whether exercise intensity prescription based on VTs is favorable to prescription based on RPE, %HRR, or %PO peak in terms of improvements in cardiorespiratory fitness. This should be further investigated in future research. Moreover, as the large random error within and between raters suggests, training schemes based on VTs should be clinically evaluated on individual level. For example, a talk test may be used to evaluate whether the intensity is either too high or too low [39]. If this appears to be the case, VT determination should be critically re-evaluated by one or more experts in order to prevent over- or undertraining in that individual. In addition, the low absolute agreement between raters suggest that during a longitudinal follow-up with several GXTs within an individual, it would be advised to identify the VTs by the same rater.

Study limitations

Although the sample size of the present study was equal to or higher than the sample size in comparable studies [17, 18, 20, 22], the sample size of the subgroups, especially for individuals with tetraplegia, was small. Therefore the statistical power was reduced, which makes interpretation of the ICCs for subgroups less reliable. It must be noted, however, that large sample sizes in rehabilitation populations are difficult to obtain. Another aspect that was not investigated in the present study is the test-retest reliability across days of the GXT itself. It might be interesting for future studies to investigate reliability of VTs during repeated GXTs, as the variability of VTs between tests within individuals is unknown for this population.

Conclusions

Ninety percent of VTs could be determined. Most of the VTs that could not be determined were VT2s and related to tests in individuals with tetraplegia. For the VTs that could be determined, the relative intrarater reliability was very high with small to large random error. The relative interrater reliability was high to very high with large random error. Although these results are positive on group level and show that determination of VTs might be a promising method to define training intensity for the majority of the tested recreationally active individuals with SCI, it should be noted that a critical evaluation of the VTs is necessary and other exercise intensity prescription methods should be considered when one or both VTs cannot be determined.

Data archiving

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments The authors would like to thank Marcel Post, Center of Excellence in Rehabilitation Medicine, De Hoogstraat Rehabilitation, Utrecht, the Netherlands, for his support with the manuscript, and Ilse Blokland, Research and Development, Heliomare Rehabilitation Center, Wijk aan Zee, the Netherlands, for her support with Matlab programming.

Funding This study was funded by the National Institutes of Health National Institute of Neurological Disorders and Stroke (grant no. NS083064), Miami Project to Cure Paralysis, HandicapNL, Stichting Mitialto, Stichting Beatrixoord Noord-Nederland, University Medical Center Groningen, Heliomare Rehabilitation Center, and Stichting Handbike Events.

Author contributions IKO was responsible for developing the idea/research question, developing the protocol for data analysis, statistical analysis, interpreting results, writing the report. REC was responsible for designing the test protocol, inclusion of participants, executing the graded exercise tests, interpreting results, providing feedback on the report. JLM was responsible for designing the test protocol, inclusion of participants, executing the graded exercise tests, interpreting results, providing feedback on the report. FPG was responsible for developing the protocol for data analysis, analysis/rating of test data, providing feedback on the report. FRI was responsible for developing the protocol for data analysis, analysis/rating of test data, providing feedback on the report. LJV was responsible for developing the idea/research question, interpreting results, providing feedback on the report. LHW was responsible for developing the idea/research question, interpreting results, providing feedback on the report. SGR was responsible for developing the idea/research question, interpreting results, providing feedback on the report.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Statement of ethics We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research.

References

1. Haisma JA, Van Der Woude LHV, Stam HJ, Bergen MP, Sluis TAR, Bussmann JBJ. Physical capacity in wheelchair-dependent persons with a spinal cord injury: a critical review of the literature. Spinal Cord. 2006;44:642–52.
2. Garshick E, Kelley A, Cohen SA, Garrison A, Tun CG, Gagnon D, et al. A prospective assessment of mortality in chronic spinal cord injury. Spinal Cord. 2005;43:408–16.
3. Nightingale TE, Metcalfe RS, Vollaard NB, Bilzon JL. Exercise guidelines to promote cardiometabolic health in spinal cord injured humans: time to raise the intensity? Arch Phys Med Rehabil. 2017;98:1693–704.
4. Valent LJM, Dallmeijer AJ, Houdijk JHP, Slootman JR, Janssen TWJ, Van Der Woude LHV. Effects of hand cycle training on wheelchair capacity during clinical rehabilitation in persons with a spinal cord injury. Disabil Rehabil. 2010;32:2191–2200.
5. Valent LJM, Dallmeijer AJ, Houdijk H, Slootman HJ, Janssen TW, Post MWM, et al. Effects of hand cycle training on physical capacity in individuals with tetraplegia: a clinical trial. Phys Ther. 2009;89:1051–60.
6. Valent LJ, Dallmeijer AJ, Houdijk H, Slootman HJ, Post MW, van der Woude LH. Influence of hand cycling on physical capacity in the rehabilitation of persons with a spinal cord injury: a longitudinal cohort study. Arch Phys Med Rehabil. 2008;89:1016–22.
7. De Groot S, Postma K, Van Vliet L, Timmermans R, Valent LJM. Mountain time trial in handcycling: Exercise intensity and predictors of race time in people with spinal cord injury. Spinal Cord. 2014;52:455–61.
8. Hoekstra S, Valent L, Gobets D, van der Woude L, de Groot S. Effects of four-month handbike training under free-living conditions on physical fitness and health in wheelchair users. Disabil Rehabil. 2017;39:1581–8.
9. Wolpern AE, Burgos DJ, Janot JM, Dalleck LC. Is a threshold-based model a superior method to the relative percent concept for establishing individual exercise intensity? A randomized controlled trial. BMC Sports Sci Med Rehabil. 2015;7:16. https://doi.org/10.1186/s13102-015-0011-z.
10. Meyer T, Gabriel HHW, Kindermann W. Is determination of exercise intensities as percentage of VO2max or HRmax adequate? Med Sci Sport Exerc. 1999;31:1342–5.
11. Kindermann W, Simon G, Keul J. The significance of the aerobic-anaerobic transition for the determination of work load intensities during endurance training. Eur J Appl Physiol. 1979;34:25–34.
12. Meyer T, Lucía A, Earnest CP, Kindermann W. A conceptual framework for performance diagnosis and training prescription from submaximal gas exchange parameters – Theory and application. Int J Sport Med. 2005;26:38–48.
13. Seiler KS, Kjerland GØ. Quantifying training intensity distribution in elite endurance athletes: Is there evidence for an "optimal" distribution? Scand J Med Sci Sport. 2006;16:49–56.
14. Mezzani A, Hamm LF, Jones AM, McBride PE, Moholdt T, Stone JA, et al. Aerobic exercise intensity assessment and prescription in cardiac rehabilitation. J Cardiopulm Rehabil Prev. 2012;32:327–50.
15. Gaskill SE, Ruby BC, AVAJ Walker, Sanchez OA, Serfass RC, Leon AS. Validity and reliability of combining three methods to determine ventilatory threshold. Med Sci Sport Exerc. 2001;33:1841–8.

16. Shimizu M, Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, et al. The ventilatory threshold: Method, protocol, and evaluator agreement. Am Heart J. 1991;122:509–16.
17. Gladden LB, Yates JW, Stremel RW, Stamford BA. Gas exchange and lactate anaerobic thresholds: inter- and intraevaluator agreement. J Appl Physiol. 1985;58:2082–9.
18. Aunola S, Rusko H. Reproducibility of aerobic and anaerobic thresholds in 20-50 year old men. Eur J Appl Physiol Occup Physiol. 1984;53:260–6.
19. Leicht CA, Griggs KE, Lavin J, Tolfrey K, Goosey-Tolfrey VL. Blood lactate and ventilatory thresholds in wheelchair athletes with tetraplegia and paraplegia. Eur J Appl Physiol. 2014;114:1635–43.
20. Coutts KD, McKenzie DC. Ventilatory thresholds during wheelchair exercise in individuals with spinal cord injuries. Paraplegia. 1995;33:419–22.
21. Au JS, Sithamparapillai A, Currie KD, Krassioukov AV, MacDonald MJ, Hicks AL. Assessing ventilatory threshold in individuals with motor-complete spinal cord injury. Arch Phys Med Rehabil. 2018;99:1991–7.
22. Bhambhani YN, Burnham RS, Wheeler GD, Eriksson P, Holland LJ, Steadward RD. Ventilatory threshold during wheelchair exercise in untrained and endurance-trained subjects with quadriplegia. Adapt Phys Act Q. 1995;12:333–43.
23. Maher JL, Cowan RE. Comparison of 1- versus 3-minute stage duration during arm ergometry in individuals with spinal cord injury. Arch Phys Med Rehabil. 2016;97:1895–900.
24. Kouwijzer I, Valent L, Osterthun R, van der Woude LHV, De Groot S, HandbikeBattle group. Peak power output in handcycling of individuals with a chronic spinal cord injury: predictive modeling, validation and reference values. Disabil Rehabil. 2018:1–10. https://doi.org/10.1080/09638288.2018.1501097. [Epub ahead of print]
25. Janssen TWJ, Van Oers CAJ, Hollander PA, Veeger DHEJ, van der Woude LHV. Isometric strength, sprint power, and aerobic power in individuals with a spinal cord injury. Med Sci Sports Exerc. 1993;25:863–70.
26. Buchfuhrer MJ, Hansen JE, Robinson TE, Sue DY, Wasserman K, Whipp BJ. Optimizing the exercise protocol for cardiopulmonary assessment. J Appl Physiol Respir Environ Exerc Physiol. 1983;55:1558–64.
27. Robergs RA, Dwyer D, Astorino T. Recommendations for improved data processing from expired gas analysis indirect calorimetry. Sport Med. 2010;40:95–111.
28. Kuipers H, Verstappen FTJ, Keizer P, Geurten P, van Kranenburg G. Variability of aerobic performance in the laboratory and its physiologic correlates. Int J Sports Med. 1985;6:197–201.
29. Wasserman K, Hansen JE, Sue DY, Stringer WW, Sietsema, KE, Sun X-G, et al. Principles of exercise testing and interpretation. 5th ed. Philadelphia, USA. Lippincott Williams & Wilkins; 2012.
30. Beaver WL, Wasserman K, Whipp BJ. A new method for detecting anaerobic threshold by gas exchange. J Appl Physiol. 1986;60:2020–7.
31. Binder RK, Wonisch M, Corra U, Cohen-Solal A, Vanhees L, Saner H, et al. Methodological approach to the first and second lactate threshold in incremental cardiopulmonary exercise testing. Eur J Cardiovasc Prev Rehabil. 2008;15:726–34.
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–10.
33. Munro BH. Statistical methods for health care research. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2004. p. 239–58.
34. Simmons OL, Kressler J, Nash MS. Reference fitness values in the untrained spinal cord injury population. Arch Phys Med Rehabil. 2014;95:2272–8.
35. Lovell D, Shields D, Beck B, Cuneo R, McLellan C. The aerobic performance of trained and untrained handcyclists with spinal cord injury. Eur J Appl Physiol. 2012;112:3431–7.
36. Schneider DA, Sedlock DA, Gass E, Gass G. $\dot{V}$O2peak and the gas-exchange anaerobic threshold during incremental arm cranking in able-bodied and paraplegic men. Eur J Appl Physiol Occup Physiol. 1999;80:292–7.
37. Valent LJM, Dallmeijer AJ, Houdijk H, Slootman J, Janssen TWJ, Hollander AP, et al. The individual relationship between heart rate and oxygen uptake in people with a tetraplegia during exercise. Spinal Cord. 2007;45:104–11.
38. van der Scheer JW, Hutchinson MJ, Paulson T, Martin Ginis KA, Goosey-Tolfrey VL. Reliability and validity of subjective measures of aerobic intensity in adults with spinal cord injury: A systematic review. PM R. 2018; 10:194–207.
39. Cowan R, Ginnity K, Kressler J, Nash M. Assessment of the talk test and rating of perceived exertion for exercise intensity prescription in persons with paraplegia. Top Spinal Cord Inj Rehabil. 2012;18:212–9.