

RESEARCH ARTICLE

# Reliability and Responsiveness of Cardiopulmonary Exercise Testing in Fatigued Persons with Multiple Sclerosis and Low to Mild Disability

Martin Heine<sup>1\*</sup>, Lizanne Eva van den Akker<sup>2</sup>, Olaf Verschuren<sup>1</sup>, Anne Visser-Meily<sup>1</sup>, Gert Kwakkel<sup>3,4</sup>, TREFAMS-ACE Study Group<sup>1</sup>

**1** Brain Center Rudolf Magnus and Center of Excellence for Rehabilitation Medicine, University Medical Center Utrecht and Rehabilitation Center De Hoogstraat, Rembrandtkade 10, 3583 TM, Utrecht, The Netherlands, **2** Department of Rehabilitation Medicine & EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands, **3** Department of Rehabilitation Medicine, MOVE Research Institute Amsterdam, VU University Medical Center, Amsterdam, The Netherlands, **4** Department of Neurorehabilitation, Centre of Rehabilitation and Rheumatology READE, Amsterdam, The Netherlands

<sup>†</sup> Membership of the TREFAMS-ACE Study Group is provided in [S1 Appendix](#).

\* [m.heine@dehoogstraat.nl](mailto:m.heine@dehoogstraat.nl)



**OPEN ACCESS**

**Citation:** Heine M, van den Akker LE, Verschuren O, Visser-Meily A, Kwakkel G, TREFAMS-ACE Study Group. (2015) Reliability and Responsiveness of Cardiopulmonary Exercise Testing in Fatigued Persons with Multiple Sclerosis and Low to Mild Disability. PLoS ONE 10(3): e0122260. doi:10.1371/journal.pone.0122260

**Academic Editor:** Bart Rypma, University of Texas at Dallas, UNITED STATES

**Received:** July 11, 2014

**Accepted:** February 15, 2015

**Published:** March 19, 2015

**Copyright:** © 2015 Heine et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The TREFAMS-ACE study is funded by the Fonds NutsOhra (ZonMw 89000005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

### Background

Peak oxygen uptake ( $VO_{2peak}$ ) via cardiopulmonary exercise testing is considered the gold standard for testing aerobic capacity in healthy participants and people with various medical conditions. The reliability and responsiveness of cardiopulmonary exercise testing outcomes in persons with MS (PwMS) have not been extensively studied.

### Objective

(1) to investigate the reliability of cardiopulmonary exercise parameters in PwMS; (2) to determine the responsiveness, in terms of the smallest detectable change (SDC), for each parameter.

### Design

Two repeated measurements of cardiopulmonary exercise outcomes were obtained, with a median time interval of 16 days.

### Methods

Thirty-two PwMS suffering from subjective fatigue performed cardiopulmonary exercise tests on a cycle ergometer, to voluntary exhaustion. We calculated the reliability, in terms of the intra-class correlation coefficient (ICC [2,k]; absolute agreement), and the measurement error, in terms of standard error of measurement (SEM) and SDC at individual ( $SDC_{individual}$ ) and group level ( $SDC_{group}$ ).

## Results

The ICC for  $\text{VO}_2\text{peak}$  was 0.951, with an SEM of  $0.131 \text{ L}\cdot\text{min}^{-1}$  and an  $\text{SDC}_{\text{individual}}$  of  $0.364 \text{ L}\cdot\text{min}^{-1}$ . When corrected for bodyweight, the ICC of  $\text{VO}_2\text{peak}$  was 0.933, with an SEM of  $1.7 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$  and in an  $\text{SDC}_{\text{individual}}$  of  $4.6 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ .

## Limitations

Generalization of our study results is restricted to fatigued PwMS with a low to mild level of disability.

## Conclusions

At individual level, cardiopulmonary exercise testing can be used reliably to assess physical fitness in terms of  $\text{VO}_2\text{peak}$ , but less so to determine significant changes. At group level,  $\text{VO}_2\text{peak}$  can be reliably used to determine physical fitness status and establish change over time.

## Introduction

Multiple sclerosis (MS) is a chronic demyelinating inflammatory disorder of the central nervous system (CNS) with unknown etiology.[1] Progressive demyelination within the CNS results in a range of motor, sensory, cerebellar and cognitive dysfunctions, many of which may, directly or indirectly, limit physical fitness. The resulting impairments and limitations in daily activities may subsequently result in considerable physical deconditioning and even in comorbidities that reduce life expectancy.[2–4] Maintaining or improving physical fitness in persons with MS (PwMS) is paramount. In the past two decades, several intervention studies have investigated the feasibility and potential benefits of various types of exercise training in PwMS.[5–8] These studies have showed that training regimes like endurance, resistance or combined training can be used safely by PwMS. In addition, these training regimes may prevent deconditioning or improve physical fitness in these patients.[5–8] Measuring physical fitness is important in the context of clinical practice to evaluate people's health status, and in the context of research to evaluate the effects of exercise regimes or other interventions.[7] Cardiopulmonary exercise testing (CPET) using maximal cycling ergometry is considered the 'gold standard' for the assessment of exercise tolerance in both healthy people and people with various medical conditions.[9] CPET is preferably accompanied by measurements of ventilation and gas exchange (i.e.  $\text{VO}_2\text{peak}$ ) for direct assessment of aerobic function.[10] If a measure is to be used in clinical practice, it should be both content-valid and reliable. In this case, consecutive measurements from a given subject who has not changed, should be the same under several conditions (test-retest).[11]

Recently, we showed that CPET is a valid measure of physical fitness in PwMS with a low to mild level of disability (Expanded Disability Status Scale;  $\text{EDSS} \leq 4.0$ ).[12] In addition, we showed that the Oxygen Uptake Efficiency Slope (OUES), as a potential alternative to  $\text{VO}_2\text{peak}$  in PwMS who are unable to attain maximal exercise intensities, is highly predictive of  $\text{VO}_2\text{peak}$ . [13] Even though the validity and the responsiveness of CPET in PwMS have not been studied extensively until recent[8, 12, 14], the CPET has been used as an outcome in clinical trials quite often already.[8, 15–20] For instance, in a study by Petajan et al., a 22% improvement in  $\text{VO}_2\text{peak}$  was found following a 15 week aerobic training program.[20] In contrast,

Mostert and Kesseling were unable to replicate these results, in terms of  $\text{VO}_2\text{peak}$ , in a shorter but more intensive training program.[19] However, little is known about the extent to which maximal exercise performance in PwMS is hampered by day-to-day variation. In other words, how confident are we that the reported 22% change in  $\text{VO}_2\text{peak}$  is not due to measurement error? To our knowledge, one study has suggested that CPET can be used reliably in PwMS under stringent laboratory conditions.[14] This study, by Langeskov-Christensen and colleagues reported that a change of ~10% is required to consider this change significant. However, MS is an unstable medical condition with symptom experience changes even within days, yet little is known on the reliability and responsiveness of CPET in a less stringent laboratory setting which may, for instance, substantially affect our confidence in the 22% change in  $\text{VO}_2\text{peak}$  as reported by Petajan et al.[20] This question is important in the light of clinical research and practice, as the responsiveness of a test may provide valuable information on the required number of participants (i.e. statistical power) in clinical trials using CPET, on the ability of the test to determine individual changes in  $\text{VO}_2\text{peak}$  in clinical practice, and to determine if the change in  $\text{VO}_2\text{peak}$  following a clinical trial can be considered beyond measurement error. Our research question was therefore: what is the reliability and responsiveness of CPET, and in particular  $\text{VO}_2\text{peak}$ , as assessed by two consecutive CPETs when applied to PwMS?

The objective of the present study was twofold: (1) to investigate the reliability of CPET, and in particular  $\text{VO}_2\text{peak}$ , in PwMS; (2) to determine the responsiveness, in terms of the smallest detectable change.

## Participants and Methods

### Participant selection

A joint convenience sample was composed from patients treated at two study centers (VU Medical Center, Amsterdam, the Netherlands and St. Antonius Medical Center, Nieuwegein, the Netherlands). The sample consisted of 32 participants of the TREFAMS-ACE study (Treating Fatigue in Multiple Sclerosis: Aerobic Therapy, Cognitive Behavioral Therapy, Energy Conservation Management; ISRCTN69520623 and ISRCTN58583714) with definite MS.[21] No a priori power analysis was conducted. The sample size of 32 was reached by requesting participation in this reliability study of all TREFAMS-ACE participants who attended the first or second follow-up measurement between August 2012 and October 2013. The TREFAMS-ACE study and the present reliability study were approved by the Medical Ethics Board of the VU University Medical Center Amsterdam. Each potential participant received both written and oral information about the TREFAMS-ACE trial before providing written informed consent. In addition, participants provided supplementary consent for participation in the present reliability study, which involved an additional exercise test. To be included in the TREFAMS-ACE study, participants had to experience substantial fatigue ( $\geq 35$  on the fatigue subscale of the Checklist on Individual Strength[22]), had to be ambulant ( $\leq 6.0$  on the Expanded Disability Status Scale [EDSS][23]) and had to have no clinical depression ( $\leq 11$  on the Depression subscale of the Hospital Anxiety and Depression Scale[24]). Potential participants were excluded if they had participated in a professionally supervised therapy program to alleviate fatigue in the three months before inclusion. Potential participants were also excluded if they had comorbidities precluding maximal exercise participation, or MS relapse confirmed by a neurologist in the three months prior to study participation. Details of the TREFAMS-ACE study have been described elsewhere.[21]

Prior to the cardiopulmonary exercise test, self-reported baseline characteristics and disease specifics were recorded: sex, age, height, weight, body mass index (BMI), time since diagnosis and type of MS (classified as relapsing-remitting, primary progressive, secondary progressive or not specified). Neurological disability was determined by a certified physician using the EDSS.[23]

## Cardiopulmonary exercise testing (CPET)

Two identical incremental exercise tests were performed on a programmable, electromagnetically braked cycle ergometer (Kettler X7, Heinz Kettler, Germany) to determine maximal exercise capacity. Within each study center, one assessor conducted the test and re-test. The two exercise tests were conducted during the follow-up phase of the TREFAMS-ACE study, but at least 10 weeks after the intervention phase. A period of one to three weeks was scheduled between the two consecutive tests, depending on the participants' flexibility in making an appointment. Participants were asked not to change their exercise habits between the two tests. During the intervention phase of the TREFAMS-ACE study, participants had already performed at least one exercise test, and were thus familiar with the test protocol. To avoid bias, the two assessors in each center and the participants were not informed of the results of the first test prior to the second test.

Handlebars and saddle of the cycle ergometer were adjusted to match each participant's anthropometrics. Following a 3-minute rest phase, participants started cycling at 25 Watt, with power output increasing by 10 Watt (women) or 15 Watt (men) each minute. This protocol was based on the rationale that small increments prolong exercise duration and enable participants with peripheral limitations to sustain longer exercise duration. Preferably, the exercise was terminated by voluntary exhaustion following 8–12 minutes of exercise.[25] During the exercise testing, participants were asked to maintain a cadence of 60–80 rotations per minute (rpm). Participants were verbally encouraged, especially beyond a respiratory exchange ratio  $> 1.00$ . The exercise test was terminated by voluntary exhaustion, by a cadence  $< 45$  rpm or for safety reasons, compliant with the American College of Sports Medicine's guidelines for clinical exercise testing.[26]

Each participant's gas exchange was measured using a portable mixing-chamber monitoring system (Cortex MetaMax 3B; Cortex Medical; Germany). The Cortex Metamax is a valid and reliable system for measuring ventilation parameters during exercise.[27, 28] Volume, pressure, and gas analyzers were calibrated automatically by the system prior to each test using a 3L syringe, atmospheric pressure at the time (mmHg), and both ambient and reference gases (4.98% CO<sub>2</sub>, 17.05% O<sub>2</sub>), respectively. Raw data was averaged over 10-second intervals for analysis. Peak exercise values were defined as the highest recorded 10-second average. The following parameters derived from the CPET were used in the present study: peak oxygen consumption (VO<sub>2peak</sub>), maximal work output (W<sub>max</sub>), maximal ventilation (VE<sub>max</sub>), maximal respiratory exchange ratio (RER), perceived exertion immediately following voluntary exhaustion, measured by the 20-point BORG scale[29], and maximal heart rate. In the current literature VO<sub>2peak</sub> is often, but not always, expressed relative to bodyweight. Hence, the present study used both the corrected and uncorrected VO<sub>2peak</sub>. In addition, we included the Oxygen Uptake Efficiency Slope (OUES). The OUES may be a valuable submaximal measure of physical fitness in those PwMS for whom VO<sub>2peak</sub> is a measure of performance, rather than aerobic capacity.[13] The OUES was determined by measuring the slope of VO<sub>2</sub> (mL·kg<sup>-1</sup>·min<sup>-1</sup>) and log<sub>10</sub> VE (L·min<sup>-1</sup>) until voluntary exhaustion. The OUES has been claimed to be independent of exercise duration, exercise protocol and assessor.[30, 31] It has also been shown that the OUES is highly predictive of VO<sub>2peak</sub> in PwMS.[13]

## Statistics

From a psychometric point of view, reliability can be defined as the proportion of total variance in the measurements which indicates a 'true' difference between patients.[32] In this concept, each observation is assumed to be composed of two components: a true score and an error associated with the association.[11] Based on this definition, the proportion of between-patient

variance compared to the total variance was calculated by the intra-class correlation coefficients (ICC [2, k]) model (i.e., a two-way random model with absolute agreement). An ICC greater than 0.80 reflects excellent reliability, whereas ICCs from 0.70 to 0.79 reflect good reliability.[33] The recommended minimum for the lower limit of the 95% confidence interval (CI) is 0.85.[34] Measurement error was defined as the systematic and random error in a patient’s score that is not attributed to true changes in CPET performance. The standard error of measurement ( $SEM_{agreement}$ ) was used for the measurement error, to determine the precision of the total score of both tests. The SEM describes the error in interpreting an individual’s test score. It allows the ‘true’ test performance to be estimated using a reliability coefficient and is computed by multiplying the standard deviation of the total score by the square root of 1 minus its reliability coefficient ( $SEM = SD_{pooled} \times \sqrt{1-ICC}$ ).[35] The smallest detectable change ( $SDC_{individual}$ ) in each aerobic capacity outcome was computed as  $1.96 \times \sqrt{2} \times SEM$  to obtain a 95% CI.[32, 35] In addition we considered it valuable, especially for clinical research, to determine the smallest detectable difference at group level ( $SDC_{group}$ ) rather than at individual level. The  $SDC_{group}$  was calculated as  $SDC_{individual} / \sqrt{n}$ . The Bland–Altman procedure was used to check for heteroscedasticity of the test and retest for each outcome measure.[36] All analyses were conducted using the SPSS statistical package (version 19.0).

## Results

A total of 13 men and 19 women (Table 1) completed both tests without complications. The median disability level on the EDSS scale was 2.5 (Interquartile range (IQR) 2.0–3.0) Most participants had relapsing-remitting MS (N = 27). The second test was conducted after a median interval of 16 days (IQR 14–19 days) following the first.

### Reliability and responsiveness

The reliability and responsiveness of the CPET outcomes are presented in Table 2. The reliability of both the %HRmax and the BORG score was based on 31 instead of 32 participants due to one missing value, as the heart rate monitor lost signal during the exercise test for one participant, and the BORG score was not acquired within one minute post voluntary exhaustion for

**Table 1. Participant characteristics.**

Participant characteristics	N = 32	SD (if mean)
		IQR (if median)
Sex (N) men women	1319	
Type of MS (N)Relapsing-Remitting Primary ProgressiveSecondary ProgressiveNot specified	27122	
Mean age, years	50.7	8.1
Mean height, cm	176	11
Mean weight, kg	79	15
Mean BMI, kg/m <sup>2</sup>	25.6	3.7
Median time since diagnosis, years	9	4–13
Median EDSS score	2.5	2.0–3.0
Median days between tests	16	14–19

BMI, Body Mass Index; EDSS, Expanded Disability Status Scale; IQR, Interquartile range; N, number of participants; SD, standard deviation.

doi:10.1371/journal.pone.0122260.t001

**Table 2. Reliability and responsiveness of CPET in 32 PwMS.**

Outcome	Test 1 Mean (SD <sub>1</sub> )	Test 2 Mean (SD <sub>2</sub> )	MD (SD <sub>pooled</sub> )	ICC	95%CI of ICC	SEM	SDC <sub>individual</sub>	SDC <sub>group</sub>	LOA
VO <sub>2</sub> peak (L·min <sup>-1</sup> )	1.745 (0.593)	1.720 (0.593)	-0.03 (0.19)	0.951	0.902 to 0.976	0.131	0.364	0.06	-0.40 to 0.35
VO <sub>2</sub> peak (mL·kg <sup>-1</sup> ·min <sup>-1</sup> )	21.7 (6.4)	21.5 (6.4)	-0.29 (2.36)	0.933	0.868 to 0.967	1.7	4.6	0.82	-5.0 to 4.3
Wmax (Watt)	157.8 (51.6)	158.8 (53.4)	0.94 (12.92)	0.971	0.941 to 0.985	9.1	25.2	4.46	-24.9 to 26.8
RER	1.19 (0.13)	1.20 (0.12)	0.01 (0.12)	0.565	0.273 to 0.761	0.08	0.23	0.04	-0.22 to 0.24
VE (L/min)	66.0 (24.7)	65.1 (25.4)	-0.81 (9.04)	0.936	0.874 to 0.968	6.4	17.7	3.13	-19 to 17
%HRmax (n = 31)	90.6 (10.8)	90.3 (9.8)	-0.22 (5.03)	0.886	0.777 to 0.943	3.5	9.8	1.73	-10.3 to 9.8
BORG (n = 31)	17.4 (1.8)	17.6 (2.0)	0.03 (1.45)	0.706	0.466 to 0.849	1.1	2.9	0.52	-2.9 to 2.9
OUES	24.1 (7.5)	23.6 (6.4)	-0.48 (3.34)	0.887	0.782 to 0.943	2.4	6.6	1.16	-7.2 to 6.2

BORG, rating of perceived exertion following voluntary exhaustion; HRmax, maximal heart rate expressed as percentage of the age-predicted maximal heart rate (%); ICC, Intra-class Correlation Coefficient including 95% Confidence Interval (95%CI); LOA, Limits Of Agreement; OUES, oxygen uptake efficiency slope; RER, respiratory exchange ratio; SD, Standard Deviation; SDC, Smallest Detectable Change; SEM, Standard Error of Measurement; VE, peak ventilation (L·min<sup>-1</sup>); VO<sub>2</sub>peak, peak oxygen uptake (L·min<sup>-1</sup> and mL·kg<sup>-1</sup>·min<sup>-1</sup>); Wmax, maximal power output (Watt).\

doi:10.1371/journal.pone.0122260.t002

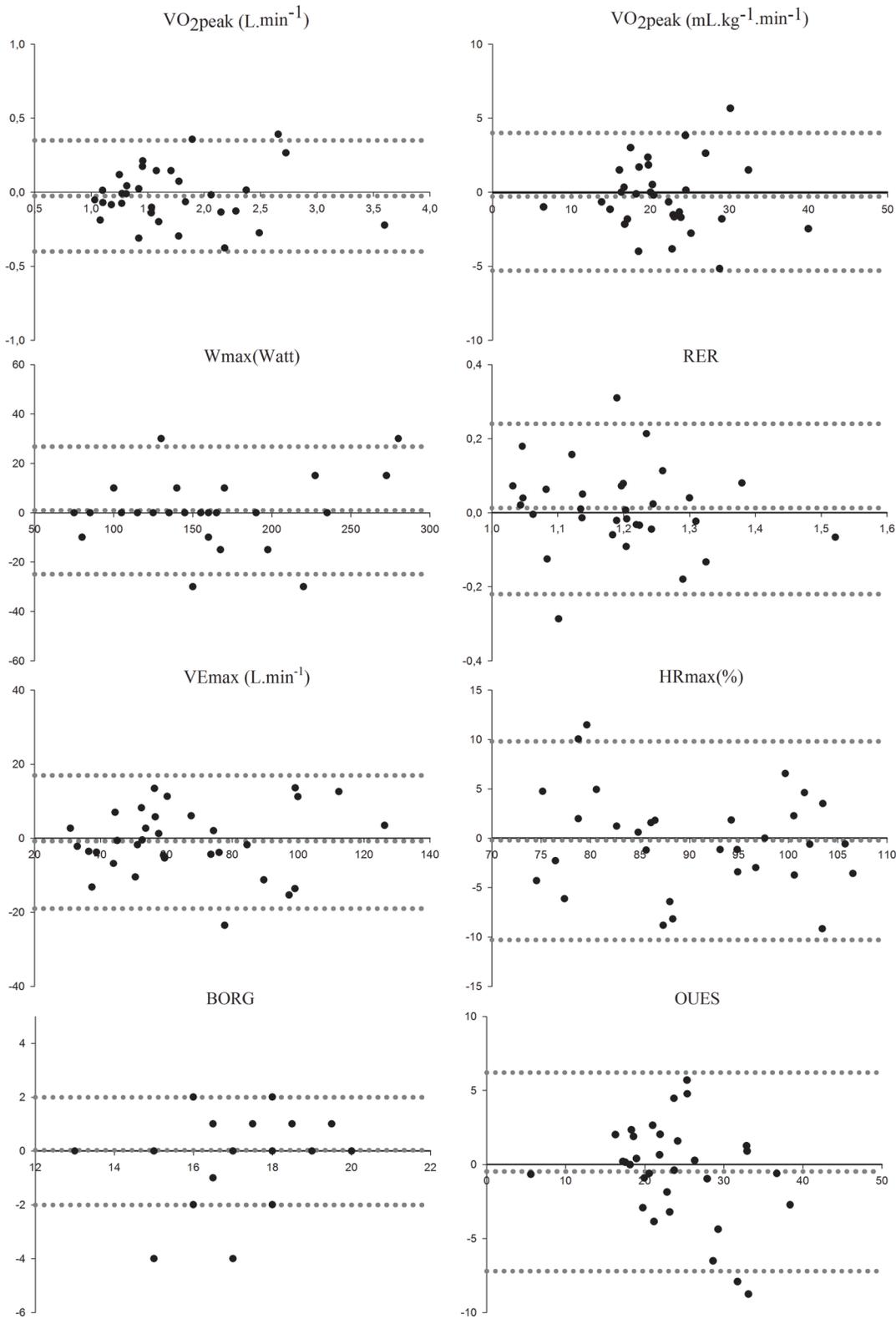
another participant. The ICC of VO<sub>2</sub>peak was 0.951, with an SEM of 0.131 L·min<sup>-1</sup> and an SDC<sub>individual</sub> of 0.364 L·min<sup>-1</sup>. After correcting for body weight, the ICC for VO<sub>2</sub>peak was 0.933, with an SEM of 1.7 mL·kg<sup>-1</sup>·min<sup>-1</sup> and an SDC<sub>individual</sub> of 4.6 mL·kg<sup>-1</sup>·min<sup>-1</sup>. The ICC for maximal power was 0.971, with an SEM of 9.1W and an SDC of 25.2W. As regards peak ventilation, the ICC was found to be 0.936, with an SEM of 6.4 L/min and an SDC<sub>individual</sub> of 17.7 L/min. As can be seen from the Bland-Altman plots (Fig. 1) there were no signs of heteroscedasticity, except for the OUES. Potential outliers were included in the analysis.

## Discussion

The purpose of the present study was twofold: (1) to investigate the reliability of cardiopulmonary exercise parameters in PwMS; (2) to determine the responsiveness, in terms of the smallest detectable change, of each parameter. The aerobic capacity, in terms of VO<sub>2</sub>peak, was low (21.6 ± 6.4 mL·kg<sup>-1</sup>·min<sup>-1</sup>) when compared to reference values for sedentary healthy people (39.0 ± 6.8 mL·kg<sup>-1</sup>·min<sup>-1</sup> for sedentary men; 30.0 ± 5.4 mL·kg<sup>-1</sup>·min<sup>-1</sup> for sedentary women). [37] The present study showed that the reliability of VO<sub>2</sub>peak, Wmax and VE could be considered excellent. However, we found that some individuals may need large changes in aerobic capacity (>21%) for them to be considered real changes beyond the 95%CI intervals of measurement error.

### Reliability and responsiveness

The excellent reliability in two consecutive tests of VO<sub>2</sub>peak, as shown by the high ICCs, is in line with ICCs previously reported for CPET. [38, 39] Langeskov-Christensen and colleagues found a Pearson correlation coefficient of 0.98, which is of the same magnitude as the ICC found in the present study, for two consecutive CPETs in a sample of PwMS with similar disability (mean EDSS = 2.6). [14] In contrast to the excellent test-retest reliability, the smallest detectable change for an individual PwMS was relatively large which questions the responsiveness of this measure. A change of at least 0.367 L·min<sup>-1</sup> or 4.6 mL·kg<sup>-1</sup>·min<sup>-1</sup> (~21%) was required for it to be considered a statistically significant improvement or deterioration. This may indicate that the incremental exercise test at individual level may not be suitable for monitoring change in aerobic capacity in PwMS. However, the small SDC<sub>group</sub> found in the



**Fig 1. Bland-Altman plots of CPET outcomes in PwMS.** BORG, rating of perceived exertion following voluntary exhaustion; HRmax, maximal heart rate expressed as percentage of the age-predicted maximal heart rate (%); OUES, oxygen uptake efficiency slope; RER, respiratory exchange ratio; VE, peak ventilation (L.min<sup>-1</sup>); VO<sub>2</sub>peak, peak oxygen uptake (L.min<sup>-1</sup> and mL.kg<sup>-1</sup>.min<sup>-1</sup>); Wmax, maximal power output (Watt).

doi:10.1371/journal.pone.0122260.g001

present study suggests that, at group level, small changes in the aerobic capacity of PwMS can be identified. The  $SDC_{\text{individual}}$  in the present study (~21%) was considerably larger than the required change found in the study by Langeskov-Christensen et al. (~10%).[\[14\]](#) One possible explanation is that this may be related to the more stringent standardization in the study by Langeskov-Christensen et al. (in terms of food intake, time between tests and exercise behavior).[\[14\]](#) A second possible explanation is that the present study used two different exercise labs, which may have introduced some additional variance. Thirdly, the present study recruited PwMS who experienced excessive fatigue. Since there was no restriction on various factors that may have influenced fatigue, perceived fatigue may have been different on each test occasion, which may explain some of the variance between the tests. However, on the contrary, the participants of the present study were familiar with the testing protocol due to the fact that this reliability study was done during the follow-up phase of an intervention study which also included CPET. Hence, this should have reduced the likelihood of a familiarization effect during the second test and reduced the between-test variance in comparison to for example Langeskov-Christensen et al.[\[14\]](#)

Recently, it was shown that  $VO_{2\text{peak}}$  is a valid outcome for physical fitness in fatigued PwMS with a low to mild level of disability.[\[12\]](#) In the present study, the mean RER and % HRmax confirm that, in PwMS and low to mild disability the CPET can be considered a valid measure of aerobic capacity.[\[12, 14\]](#) However, the outcome of CPET in PwMS with a moderate level of disability may be more closely related to function than to physical fitness.[\[12\]](#) We have suggested the oxygen uptake efficiency slope as an alternative measure of physical fitness, which can also be used in case of submaximal exercise duration.[\[13\]](#) An ICC of 0.93 has been previously reported for the OUES in healthy people.[\[40\]](#) However, the present study suggests a relatively high day-to-day variation in oxygen uptake efficiency in PwMS. A change of ~27% is needed, based on the results of the present study, for it to be considered a significant change. Alternatively, the OUES may complement the assessment of  $VO_{2\text{peak}}$ , as it may capture some of the physiological mechanisms (e.g. respiratory muscle work) that can explain the variation in  $VO_{2\text{peak}}$  between two tests.

## Limitations

Some limitations to this study need to be addressed, which may affect the interpretation of our results and can provide clues for ways to reduce measurement error and improve reliability. First, two different exercise labs were involved, with a different assessor, use of equipment and recruitment area. Research on the interrater reliability and interlab reliability of CPET in PwMS will provide additional insights into modifiable factors to reduce measurement error. Second, some of the day-to-day variation between the tests may be related to uncontrolled sources of variance, like food and beverage intake, but also to daily differences in subjective symptoms like fatigue. Third, participants of the present study had a low to mild (EDSS < 4.0) level of disability, and experienced excessive fatigue, which restricts the generalization of our study results.

## Conclusions

At individual level, CPET can be used reliably to assess physical fitness status, but less so to determine significant changes, in terms of aerobic capacity unless the test circumstances are rigorously (e.g. timing, temperature, food and beverage intake etc.) controlled between tests. This may also include subjective symptoms like fatigue. At group level, CPET can be reliably used to determine physical fitness status and establish change. Further research is warranted on the

cardiovascular working mechanisms of exercise therapy in MS, the trainability of PwMS, and clinical relevance of changes in aerobic capacity in PwMS.

## Supporting Information

### S1 Dataset. Data underlying the presented results.

(XLS)

### S1 Appendix. The TREFAMS-ACE study group.

(DOCX)

## Acknowledgments

The authors would like to thank the staff of the St. Antonius Hospital Multiple Sclerosis Center, Nieuwegein and VU University Medical Center, Amsterdam for their efforts to find eligible participants for the present study. This study was performed on behalf of the Treating Fatigue in Multiple Sclerosis: Aerobic Training, Cognitive Behavioural Therapy, Energy Conservation management (TREFAMS-ACE) Study Group ([S1 Appendix](#)).

## Author Contributions

Conceived and designed the experiments: MH GK OV AV. Performed the experiments: MH LA. Analyzed the data: MH LA OV AV GK. Wrote the paper: MH LA OV AV GK.

## References

1. Compston A, Coles A. Multiple sclerosis. *Lancet*. 2008; 372:1502–1517. doi: [10.1016/S0140-6736\(08\)61620-7](https://doi.org/10.1016/S0140-6736(08)61620-7) PMID: [18970977](https://pubmed.ncbi.nlm.nih.gov/18970977/)
2. Marrie RA, Hanwell H. General health issues in multiple sclerosis: comorbidities, secondary conditions, and health behaviors. *Continuum (Minneap Minn)*. 2013; 19:1046–1057. doi: [10.1212/01.CON.0000433284.07844.6b](https://doi.org/10.1212/01.CON.0000433284.07844.6b) PMID: [23917100](https://pubmed.ncbi.nlm.nih.gov/23917100/)
3. Marrie RA, Horwitz RI. Emerging effects of comorbidities on multiple sclerosis. *Lancet Neurol*. 2010; 9:820–828. doi: [10.1016/S1474-4422\(10\)70135-6](https://doi.org/10.1016/S1474-4422(10)70135-6) PMID: [20650403](https://pubmed.ncbi.nlm.nih.gov/20650403/)
4. Wens I, Dalgas U, Stenager E, Eijnde BO. Risk factors related to cardiovascular diseases and the metabolic syndrome in multiple sclerosis—a systematic review. *Mult Scler*. 2013; 19:1556–1564. doi: [10.1177/1352458513504252](https://doi.org/10.1177/1352458513504252) PMID: [24048545](https://pubmed.ncbi.nlm.nih.gov/24048545/)
5. Latimer-Cheung AE, Pilutti LA, Hicks AL, Martin Ginis KA, Fenuta AM, MacKibbin KA, et al. Effects of exercise training on fitness, mobility, fatigue, and health-related quality of life among adults with multiple sclerosis: a systematic review to inform guideline development. *Arch Phys Med Rehabil*. 2013; 94:1800–1828. doi: [10.1016/j.apmr.2013.04.020](https://doi.org/10.1016/j.apmr.2013.04.020) PMID: [23669008](https://pubmed.ncbi.nlm.nih.gov/23669008/)
6. Rietberg MB, Brooks D, Uitdehaag BM, Kwakkel G. Exercise therapy for multiple sclerosis. *Cochrane Database Syst Rev*. 2005:CD003980.
7. Motl RW, Pilutti LA. The benefits of exercise training in multiple sclerosis. *Nat Rev Neurol*. 2012; 8:487–497. doi: [10.1038/nrneurol.2012.136](https://doi.org/10.1038/nrneurol.2012.136) PMID: [22825702](https://pubmed.ncbi.nlm.nih.gov/22825702/)
8. Langeskov-Christensen M, Heine M, Kwakkel G, Dalgas U. Aerobic Capacity in Persons with Multiple Sclerosis: A Systematic Review and Meta-Analysis. *Sports Med*. 2015:In press.
9. Stickland MK, Butcher SJ, Marciniuk DD, Bhutani M. Assessing exercise limitation using cardiopulmonary exercise testing. *Pulm Med*. 2012; 2012:824091. doi: [10.1155/2012/824091](https://doi.org/10.1155/2012/824091) PMID: [23213518](https://pubmed.ncbi.nlm.nih.gov/23213518/)
10. Motl RW, Fernhall B. Accurate prediction of cardiorespiratory fitness using cycle ergometry in minimally disabled persons with relapsing-remitting multiple sclerosis. *Arch Phys Med Rehabil*. 2012; 93:490–495. doi: [10.1016/j.apmr.2011.08.025](https://doi.org/10.1016/j.apmr.2011.08.025) PMID: [22225573](https://pubmed.ncbi.nlm.nih.gov/22225573/)
11. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010; 63:737–745. doi: [10.1016/j.jclinepi.2010.02.006](https://doi.org/10.1016/j.jclinepi.2010.02.006) PMID: [20494804](https://pubmed.ncbi.nlm.nih.gov/20494804/)

12. Heine M, Hoogervorst EL, Hacking HG, Verschuren O, Kwakkel G. Validity of Maximal Exercise Testing in Patients With Low to Moderate Multiple Sclerosis. *Phys Ther*. 2014; 94:1168–1175. doi: [10.2522/ptj.20130418](https://doi.org/10.2522/ptj.20130418) PMID: [24677255](https://pubmed.ncbi.nlm.nih.gov/24677255/)
13. Heine M, Verschuren O, Kwakkel G. Validity of Oxygen Uptake Efficiency Slope in Patients with Multiple Sclerosis. *J Rehabil Med*. 2014; 46:656–661. doi: [10.2340/16501977-1825](https://doi.org/10.2340/16501977-1825) PMID: [24850225](https://pubmed.ncbi.nlm.nih.gov/24850225/)
14. Langeskov-Christensen M, Langeskov-Christensen D, Overgaard K, Møller AB, Dalgas U. Validity and Reliability of VO<sub>2</sub>-max Measurements in Persons with Multiple Sclerosis. *J Neurol Sci*. 2014; 342:79–87. doi: [10.1016/j.jns.2014.04.028](https://doi.org/10.1016/j.jns.2014.04.028) PMID: [24825731](https://pubmed.ncbi.nlm.nih.gov/24825731/)
15. Golzari Z, Shabkhiz F, Soudi S, Kordi MR, Hashemi SM. Combined exercise training reduces IFN-gamma and IL-17 levels in the plasma and the supernatant of peripheral blood mononuclear cells in women with multiple sclerosis. *International immunopharmacology*. 2010; 10:1415–1419. doi: [10.1016/j.intimp.2010.08.008](https://doi.org/10.1016/j.intimp.2010.08.008) PMID: [20797460](https://pubmed.ncbi.nlm.nih.gov/20797460/)
16. Rampello A, Franceschini M, Piepoli M, Antenucci R, Lenti G, Olivieri D, et al. Effect of aerobic training on walking capacity and maximal exercise tolerance in patients with multiple sclerosis: a randomized crossover controlled study. *Phys Ther*. 2007; 87:545–555. PMID: [17405806](https://pubmed.ncbi.nlm.nih.gov/17405806/)
17. Rasova K, Havrdova E, Brandejsky P, Zalisova M, Foubikova B, Martinkova P. Comparison of the influence of different rehabilitation programmes on clinical, spirometric and spiroergometric parameters in patients with multiple sclerosis. *Mult Scler*. 2006; 12:227–234. PMID: [16629428](https://pubmed.ncbi.nlm.nih.gov/16629428/)
18. Schulz KH, Gold SM, Witte J, Bartsch K, Lang UE, Hellweg R, et al. Impact of aerobic training on immune-endocrine parameters, neurotrophic factors, quality of life and coordinative function in multiple sclerosis. *Journal of the neurological sciences*. 2004; 225:11–18. PMID: [15465080](https://pubmed.ncbi.nlm.nih.gov/15465080/)
19. Mostert S, Kesselring J. Effects of a short-term exercise training program on aerobic fitness, fatigue, health perception and activity level of subjects with multiple sclerosis. *Mult Scler*. 2002; 8:161–168. PMID: [11990874](https://pubmed.ncbi.nlm.nih.gov/11990874/)
20. Petajan JH, Gappmaier E, White AT, Spencer MK, Mino L, Hicks RW. Impact of aerobic training on fitness and quality of life in multiple sclerosis. *Annals of neurology*. 1996; 39:432–441. PMID: [8619521](https://pubmed.ncbi.nlm.nih.gov/8619521/)
21. Beckerman H, Blikman LJ, Heine M, Malekzadeh A, Teunissen CE, Bussmann JB, et al. The effectiveness of aerobic training, cognitive behavioural therapy, and energy conservation management in treating MS-related fatigue: the design of the TREFAMS-ACE programme. *Trials*. 2013; 14:250. doi: [10.1186/1745-6215-14-250](https://doi.org/10.1186/1745-6215-14-250) PMID: [23938046](https://pubmed.ncbi.nlm.nih.gov/23938046/)
22. Vercoulen JH, Hommes OR, Swanink CM, Jongen PJ, Fennis JF, Galama JM, et al. The measurement of fatigue in patients with multiple sclerosis. A multidimensional comparison with patients with chronic fatigue syndrome and healthy subjects. *Arch Neurol*. 1996; 53:642–649. PMID: [8929171](https://pubmed.ncbi.nlm.nih.gov/8929171/)
23. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983; 33:1444–1452. PMID: [6685237](https://pubmed.ncbi.nlm.nih.gov/6685237/)
24. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983; 67:361–370. PMID: [6880820](https://pubmed.ncbi.nlm.nih.gov/6880820/)
25. 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; 55:1558–1564. PMID: [6643191](https://pubmed.ncbi.nlm.nih.gov/6643191/)
26. Medicine ACoS. Clinical Exercise Testing. In: Thompson WR, Gordon NF, Pescatello IS, editors. *ACSM's Guidelines for Exercise Testing and Prescription*. Philadelphia: Lippincott Williams & Wilkins; 2000. pp. 105–134.
27. Medbo JI, Mamen A, Welde B, von Heimburg E, Stokke R. Examination of the Metamax I and II oxygen analysers during exercise studies in the laboratory. *Scand J Clin Lab Invest*. 2002; 62:585–598. PMID: [12564617](https://pubmed.ncbi.nlm.nih.gov/12564617/)
28. Brehm MA, Harlaar J, Groepenhof H. Validation of the portable VmaxST system for oxygen-uptake measurement. *Gait Posture*. 2004; 20:67–73. PMID: [15196523](https://pubmed.ncbi.nlm.nih.gov/15196523/)
29. Borg G. [Physical training. 3. Perceived exertion in physical work]. *Lakartidningen*. 1970; 67:4548–4557. PMID: [5477775](https://pubmed.ncbi.nlm.nih.gov/5477775/)
30. Baba R, Nagashima M, Goto M, Nagano Y, Yokota M, Tauchi N, et al. Oxygen uptake efficiency slope: a new index of cardiorespiratory functional reserve derived from the relation between oxygen uptake and minute ventilation during incremental exercise. *J Am Coll Cardiol*. 1996; 28:1567–1572. PMID: [8917273](https://pubmed.ncbi.nlm.nih.gov/8917273/)
31. Akkerman M, van Brussel M, Hulzebos E, Vanhees L, Helders PJ, Takken T. The oxygen uptake efficiency slope: what do we know? *J Cardiopulm Rehabil Prev*. 2010; 30:357–373. doi: [10.1097/HCR.0b013e3181ebf316](https://doi.org/10.1097/HCR.0b013e3181ebf316) PMID: [20724931](https://pubmed.ncbi.nlm.nih.gov/20724931/)
32. De Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006; 59:1033–1039. PMID: [16980142](https://pubmed.ncbi.nlm.nih.gov/16980142/)

33. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*. East Norwalk, CT: Appleton & Lange; 2000.
34. Streiner D, Norman G. *Health Measurement Scales. A Practical Guide to Their Development and Use*. 2nd ed. Oxford, United Kingdom: Oxford University Press; 2000.
35. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001; 10:571–578. PMID: [11822790](#)
36. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1:307–310. PMID: [2868172](#)
37. Herdy AH, Uhlendorf D. Reference values for cardiopulmonary exercise testing for sedentary and active men and women. *Arq Bras Cardiol*. 2011; 96:54–59. PMID: [21109909](#)
38. Meyer T, Georg T, Becker C, Kindermann W. Reliability of gas exchange measurements from two different spirometry systems. *Int J Sports Med*. 2001; 22:593–597. PMID: [11719895](#)
39. 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; 4:445–453. doi: [10.1177/2047487313518474](#) PMID: [24398370](#)
40. Van Laethem C, De Sutter J, Peersman W, Calders P. Intratest reliability and test-retest reproducibility of the oxygen uptake efficiency slope in healthy participants. *Eur J Cardiovasc Prev Rehabil*. 2009; 16:493–498. doi: [10.1097/HJR.0b013e32832c88a8](#) PMID: [19483619](#)