

Test–retest reliability of skeletal muscle oxygenation measurement using near-infrared spectroscopy during exercise in patients with sport-related iliac artery flow limitation

Martijn van Hooff^{1,2}  | Eduard J. Meijer³  | Marc R. M. Scheltinga^{3,4}  |
Hans H. C. M. Savelberg²  | Goof Schep¹ 

¹Department of Sports and Exercise, Máxima Medical Centre, Veldhoven, Noord-Brabant, The Netherlands

²Department of Movement and Nutrition Sciences, Faculty of Health, Medicine and Life Sciences, Maastricht University Maastricht, Maastricht, The Netherlands

³Department of Clinical Physics, Máxima Medical Centre, Veldhoven, Noord-Brabant, The Netherlands

⁴Department of Vascular Surgery, Máxima Medical Centre, Veldhoven, Noord-Brabant, The Netherlands

Correspondence

Martijn van Hooff, Department of Sports and Exercise, Máxima Medical Centre, De Run 4600, P.O. Box 7777, 5500 MB, Veldhoven, The Netherlands.
Email: m.vanhooff@mmc.nl

Funding information

ZonMw, Grant/Award Number: 75020017

Abstract

The ankle-brachial index is an accurate tool for detecting claudication in atherosclerotic patients. However, this technique fails to identify subtle flow limitations of the iliac arteries (FLIA) in endurance athletes. Near-infrared spectroscopy (NIRS) is a noninvasive technique that measures skeletal muscle tissue oxygenation status. The aim of the present study is to examine the absolute and relative test–retest reliability of NIRS and evaluate its potential as a diagnostic tool in FLIA. NIRS-derived exercise variables were analyzed during exercise and recovery in FLIA 17 patients and 19 healthy controls. The relative reliability of absolute variables (such as the maximal value) were slight to yet predominantly substantial (intraclass correlation coefficient [ICC], ICC range: 0.06–0.76) with good to excellent absolute reliability (absolute limits of agreement [ALoA], ALoA range: 0.8 ± 10.2 to 0.7 ± 13.1 ; coefficient of variation [CV], CV range: 5%–11%). Absolute values encompassing signal amplitudes showed moderate to almost perfect relative reliability (ICC range: 0.51–0.89) and poor to good absolute reliability (ALoA range: -1.3 ± 7.0 to -2.5 ± 15.7 ; CV range: 15%–32%). Kinetic variables showed moderate to almost perfect relative reliability for most recovery kinetics variables (ICC range: 0.54–0.86) with fair to good absolute reliability (ALoA range: 0.4 ± 12.2 to 3.9 ± 37.9 ; CV range: 18%–27%). Particularly, kinetic variables showed significant differences between patients and healthy subjects. NIRS is found to be a reliable method for examining muscle tissue oxygenation variables. Given the significant differences in especially recovery kinetics between normal subjects and patients, NIRS may contribute to diagnosing FLIA in endurance athletes.

KEYWORDS

endofibrosis, maximal exercise testing, muscle oxygen saturation, near-infrared spectroscopy, reproducibility, test–retest reliability

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Clinical Physiology and Functional Imaging* published by John Wiley & Sons Ltd on behalf of Scandinavian Society of Clinical Physiology and Nuclear Medicine.

1 | INTRODUCTION

When cycling over 40 km per hour, leg pain, powerlessness, or cramping may seem like a normal exertional discomfort. However, these claudication-like symptoms may be associated with a flow limitation in the iliac artery (FLIA). Endurance athletes including cyclists and speedskaters are at risk of developing FLIA (INSITE Collaborators, 2016; Schep, Schmikli, et al., 2002). FLIA may be due to excessive arterial lengthening or kinking with hip flexion, or following intravascular lesions (endofibrosis) or a combination.

Abraham et al. reported high detection rates (90% sensitivity and 87% specificity) of a lowered ankle-brachial index (ABI) after exercise testing in patients with FLIA having moderate intravascular lesions (Abraham et al., 1997; Chevalier et al., 1986). In the absence of intimal endofibrosis, however, diagnosing flow limitations due to kinking is a challenge (INSITE Collaborators, 2016; Khan et al., 2018; Schep, Bender, et al., 2001; Sche, Kaandorp, et al., 2001; Schep, Bender, Schmikli, et al., 2002; Schep, Bender, van de Tempel, et al., 2002; Schep, Schmikli, et al., 2002). Unfortunately, no single test is sensitive and specific in all stages of severity of this disease (Schep, Bender, Schmikli, et al., 2002). Our group found that the best single test is a maximal cycling exercise test followed by ABI_{flexed} in competitive posture. In the rare case that FLIA is unilateral, the sensitivity is 73% and when bilateral the sensitivity is only 43%.

This lack of diagnostic accuracy can lead to protracted doctors' delay with unnecessary diagnostic testing and ineffective treatments (Peach et al., 2012). Therefore, there is an urgent need for a new reliable tool to accurately detect early stages of kinking and more severe abnormalities leading to FLIA in endurance athletes.

Near-infrared spectroscopy (NIRS) is a noninvasive technique that measures local muscle oxygenation indirectly during and after incremental maximal exercise, which is related to arterial muscle blood flow. It is hypothesized that patients with FLIA exhibit slower reoxygenation rates at the onset of recovery when arterial blood flow is pathologically constrained and oxygen delivery to the working muscle is impaired. Our recent proof of concept study indeed showed a characteristic NIRS response with delayed recovery kinetics. Interestingly, even subtle FLIA may be detected (van Hooff et al., 2018).

However, it is not clear which NIRS variables during which phase of the exercise test are most reliable and most discriminative for FLIA. Therefore, an in-depth exploratory study is required to compare relevant NIRS variables of healthy athletes and FLIA patients and to assess test-retest reliability of NIRS during the different phases of a standardized ramp incremental cycling test. The absolute test-retest reliability refers to the similarity of repeated measures, whereas the relative reliability reflects the ability to assess differences between subjects while taking the measurement error into account (Atkinson & Nevill, 1998; de Vet et al., 2006; Weir, 2005). The main goal of this study is to examine the test-retest reliability of exercise values (e.g., minimum or maximum), changes in signal amplitude and recovery kinetics during a maximal effort cycling test.

A perspective on differences that are observed between patients with FLIA and healthy athletes is provided.

2 | METHODS

2.1 | Subjects

The study was conducted in Máxima Medical Centre, Veldhoven, The Netherlands. Male and female subjects aged ≥ 18 years who were diagnosed with FLIA by the algorithm of Schep et al. were eligible for the study group (Schep, Bender, Schmikli, et al., 2002; Schep, Bender, van de Tempel, et al., 2002; Schep, Schmikli, et al., 2002). Exclusion criteria were earlier vascular iliac surgery, microvascular abnormalities (e.g., diabetes), vascular abnormalities outside of the iliac region, heart failure (New York Heart Association class $> I$), orthopaedic/neurological entities potentially limiting exercise capacity and obesity. These excluding conditions were considered as medical safety precautions to maximal exercise or as risk of unexpected pathophysiological effects confounding our primary outcome measures. It is known that a high level of adipose tissue thickness (ATT) influences the accuracy of NIRS measurement of underlying muscular tissue. A > 7.5 mm ATT cut-off point at the site of NIRS measurement determined with a skinfold calliper (Harpending, Batty International) was chosen. The ATT was calculated as half the skinfold thickness (Cui et al., 1991; Niemeijer et al., 2017; Van Beekvelt et al., 2001).

Healthy subjects were recruited from local cycling clubs. They completed a standardized questionnaire investigating the presence of risk factors such as smoking and positive cardiovascular family history. Candidates with FLIA were excluded according to the previously reported algorithm (Schep, Bender, Schmikli, et al., 2002; Schep, Bender, van de Tempel, et al., 2002; Schep, Schmikli, et al., 2002). Candidates who fulfilled all study criteria served as the control group.

The study protocol was approved by the local Research Ethics Committee of Maxima Medical Centre, Veldhoven, The Netherlands. The study is registered in the Dutch trial register (<https://www.trialregister.nl> identifier: Trial NL8557). The study was conducted according to the Helsinki Declaration of 1964 (Rickham, 1964). All participants gave written informed consent.

2.2 | Incremental maximal exercise tests

All study participants performed a ramp incremental cycling test twice within 4 weeks. They were instructed to abstain from strenuous exercise 48 h before the test. Dietary instructions were to refrain from a heavy meal and caffeine within 4 h before the test. Subjects performed the exercise test in competitive cycling posture on an electromagnetically braked cycle ergometer (Excalibur Sport, Lode). They were blinded to workload, time and heart rate during both sessions. Adjustments to the seating position were saved automatically in the software of the cycle ergometer to replicate identical conditions during the second test. They were instructed to

maintain a subjectively preferred constant pedalling frequency between 80 and 100 revolutions per minute. The protocol consisted of 4 min pedalling at 10% of their approximated maximal workload based on weight and physical fitness, followed by an individualized incremental maximal exercise test that was aimed to last between 8 and 12 min (O'Connor, 2013). This exercise phase was terminated when the person was not able to maintain the required preferred pedalling frequency once dropped below 70 revolutions per minute. The last registered workload was defined as the peak workload (W_{max}). Subjects were then instructed to detach their shoes from the fixated pedal connection as fast as possible and to position their feet on a resting platform that was placed over the bike (Figure 1). This manoeuvre is required as kinking of the artery resulting in FLIA is provoked by hip flexion such as in the cycling posture. Therefore, this position needs to be maintained during the recovery period, which lasted for a minimum of five minutes. During this recovery phase, automatic blood pressure measurements (Critikon 1846-SX, Soma

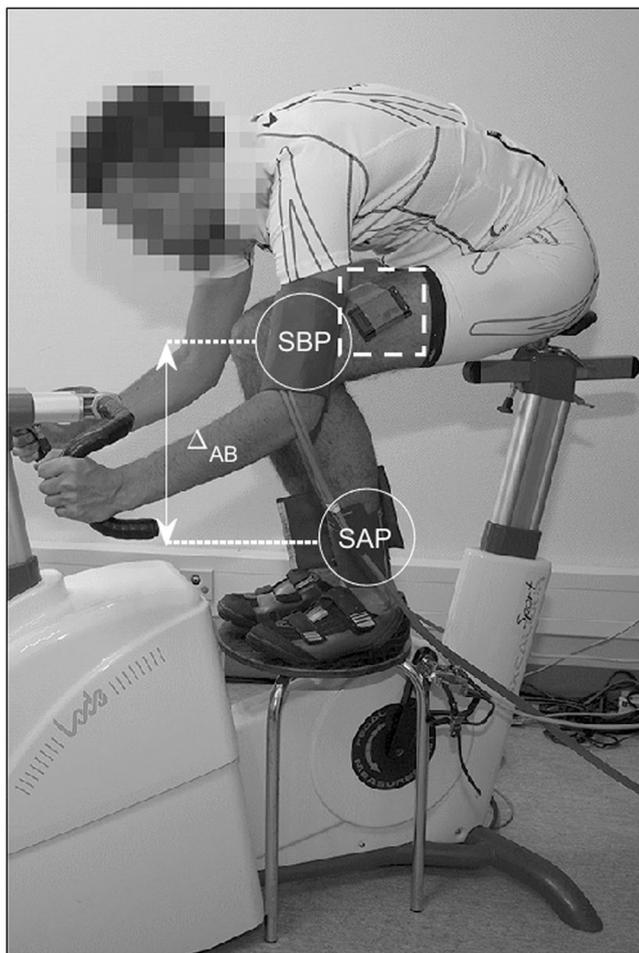


FIGURE 1 Posture during measuring the ABI and recovery kinetics using NIRS (the white circles indicate the pressure cuffs while the dashed rectangle represents the NIRS devices, black cloth are not shown in this picture to expose the devices) (van Hooff et al., 2018). ABI_{flexed}, ankle-brachial index; NIRS, near-infrared spectroscopy; SAP, systolic ankle pressure; SBP, systolic brachial pressure

Technology, Highland park Dr., Bloomfield; Duo, Datascope Corp., Mahwah, USA) were simultaneously performed in this competitive posture (trunk in an almost horizontal position). A modified ABI, ABI_{flexed}, corrected for height difference relative to the heart, is calculated as follows:

$$ABI_{flexed} = \frac{SAP - (\Delta_{AB} \times 0.78)}{SBP}$$

SAP is systolic ankle pressure (mmHg), Δ_{AB} is the vertical distance between ankle and arm (1 cm hydrostatic pressure = 0.78 mmHg) and SBP is systolic brachial pressure (mmHg). Earlier studies have demonstrated that normal values of ABI_{flexed} in healthy athletes are >0.54 whereas SAP pressure differences between legs are <23 mmHg after exercise (Gornik et al., 2008; Schep, Bender, Schmikli, et al., 2002; Schep, Bender, van de Tempel, et al., 2002; Schep, Schmikli, et al., 2002).

2.3 | NIRS

The working principle of NIRS is based on the relative transparency of biological tissue to light in the near-infrared spectrum (700–1300 nm) and on the oxygenation-dependent light-absorbing characteristics of [O₂Hb] and deoxygenated haemoglobin (HHb) concentrations in the probed tissue are computed from the different absorption properties of [O₂Hb] and (HHb) for different wavelengths of light. It is known that near-infrared light is absorbed by both haemoglobin and myoglobin. For convenience, both will be referred to as haemoglobin as their absorption cannot be distinguished. A recent extension to this type of measurement is spatially resolved spectroscopy. This technique allows the device to calculate the tissue oxygen saturation using photon diffusion (tissue saturation index, TSI (Patterson et al., 1989)). This TSI ($TSI = [O_2Hb] / ([O_2Hb] + [HHb]) \times 100$) reflects the dynamic balance between delivery and utilization of oxygen in tissue. In addition, the differential haemoglobin ($[dHb] = [O_2Hb] - [HHb]$) gives an indication of the net haemoglobin oxygenation status irrespective of changes in blood volume (Grassi et al., 1999).

NIRS measurements were performed using two wireless continuous wave NIRS (Portamon, Artinis, Elst, The Netherlands) devices, with three pairs of light-emitting diodes that emit light at two wavelengths (760 and 850 nm) that is detected by a photo-detector diode at three different source-detector distances (30, 35 and 40 mm). Data were recorded at a sample rate of 10 Hz. A differential pathlength factor of four was used as recommended by the manufacturer. In this study, we focused on the TSI, [O₂Hb] and differential haemoglobin [dHb]

Before use, the NIRS devices were wrapped in cling foil preventing moisture entering the device. The NIRS-devices were positioned on both legs and fixed on the distal muscle belly of the Vastus Lateralis (VL) muscle, 15 cm proximal to the patellar bone (Ferreira et al., 2005, 2007). A black fabric cover with elastic Velcro was placed

over the device to minimize movement and to prevent the influence of ambient light.

3 | DATA ANALYSIS

A custom-made MatLab program was used for the analysis of TSI, (O_2Hb) and $[dHb]$ data (9.6.0.1072779 (R2019a), Mathworks). Signals reflecting $[O_2Hb]$ and $[dHb]$ that are measured as changes in relative concentration were set to zero at the start of the warm-up phase. First, outliers (e.g., from unwanted body movements) were detected using a Hampel filter. This filter replaces any outlying value with the median in case this value exceeds three standard deviations from the median itself and the three neighbouring data points (Pearson, 2002). The data were subsequently filtered using a fifth-order Butterworth lowpass filter with a cutoff frequency of 1 Hz. To prevent phase shifting, zero-lag filtering was used. The endpoint of the cyclic pattern caused by pedalling in the measured signals was considered as the beginning of the recovery phase.

3.1 | Absolute values

During exercise, TSI, $[O_2Hb]$ and $[dHb]$ progressively decrease to a minimum value or arrive at a minimum plateau near-maximal exercise. Following exercise, they increase rapidly during postexercise reactive hyperemia to a maximal value. Minimal and maximal values are calculated as 5 s averages around the nadir and peak values during and after maximal exercise, respectively. The absolute difference between the minimum and maximum values is taken as the recovery amplitude (Δ_{Recovery}). Regarding TSI, the resting value is the last 10 s average before the beginning of the warmup phase. A visual explanation of these values is given in Figure 2.

3.2 | Reoxygenation kinetics

During recovery, the half-value time (HVT) was defined as the time taken to reach half the recovery amplitude during the recovery phase. In the absence of a physiologically validated kinetic model in our patient population, an empirical monoexponential curve was fit using the least-squares method on all data using the following equation:

$$Y(t) = Y_{\text{minimum}} + Y_{\Delta_{\text{Recovery}}} \times \left(1 - e^{-\frac{t}{\tau}}\right)$$

Here, Y_{minimum} is the minimum value, $Y_{\Delta_{\text{Recovery}}}$ stands for the amplitude of the recovery, whereas τ is the empirical time constant associated with the recovery. To fit such a curve, the start- and endpoints are necessary. Especially in patients, the monoexponential increase of reoxygenation does not start immediately after ceasing exercise but has a certain time delay (T_d). Therefore, an automatic method was used to detect the start (T_d) and endpoint of the monoexponential curve to avoid observer bias. This was done by using a matched filter cross-correlation method (Niemeijer et al., 2017; van Hooff et al., 2018). The moment when the correlation reached a maximal value was considered the starting point of the monoexponential curve (seen in Figure 2). The endpoint of the monoexponential curve was found by determining the highest value that was not succeeded by a higher value within 30 s during the recovery phase. This time frame with an addition of these 30 s was considered the time window for the fitting procedure. Finally, the mean response time (MRT) was calculated as the sum of the T_d and the τ . The goodness of fit (R^2 ; coefficient of determination) was considered satisfactory when exceeding 0.85 (de Groote et al., 1996; Kemps et al., 2007).

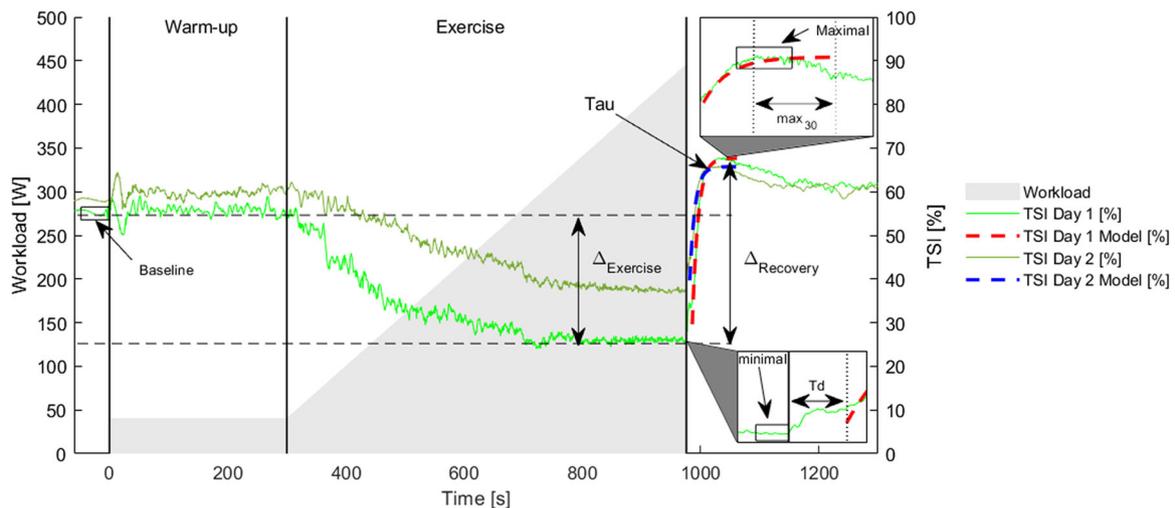


FIGURE 2 Visual presentation of TSI on during both days with explanation of measured and calculated values. The dashed lines in the recovery phase represent the monoexponential model. TSI, Tissue Saturation Index

3.3 | Statistical analysis

Statistical analysis was performed using the computing environment R (v4.04, R Development Core Team) (Team, 2021). The normality was assessed by the Shapiro–Wilk test. Results are presented as the mean value with the corresponding standard deviation. Significant differences between the two tests were determined using the paired *t* test for normal distribution and otherwise, a Wilcoxon signed-rank test. Differences between patients and healthy subjects were tested using the unpaired *t* test for normal distribution or Mann–Whitey *U* if the assumption of normal distribution was violated. The relative test–retest reliability was quantified with the intraclass correlation coefficient ($ICC_{3,1}$) coefficient and was calculated over the total group. The ICC values were interpreted following the criteria of Landis and Koch: <0.00 poor, 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1.00 almost perfect reliability (Cicchetti, 1994). The absolute reliability was determined by the absolute limits of agreement (ALoA) and visualized by Bland–Altman plots (Bland & Altman, 1986). Coefficient of variation (CV) is unit-independent reliability and consistency statistical method between two tests (or more) moments that characterize the performance of within-subject evaluation (Atkinson & Nevill, 1998; Jones & Payne, 1997; Shechtman, 2013; Synek, 2008). A $CV < 10\%$ was considered excellent, $10\% \leq CV \leq 20\%$ good, $20\% \geq CV \leq 30\%$ fair and $\geq 30\%$ poor reliability (Rosner, 2016). Correlation tests were performed using two-sided Pearson correlation coefficients. Heteroscedasticity was assessed by calculating the correlation between the absolute difference of two measurements and the mean of the two measurements. If a significant positive correlation between the absolute difference and mean value between testing days was found, ratio limits of agreements (RLoA) were calculated on ratio-scaled variables (Bland & Altman, 1999). To obtain the RLoA, the natural logarithm for both measurements with subsequent antilog of the results were calculated. Again, normal distribution was tested and checked on significant differences. The correlation of the transformed data was examined again. If a decrease in correlation was observed, these limits were considered more representative. A $p < 0.05$ was considered statistically significant for all tests.

4 | RESULTS

4.1 | Subjects

A total of 19 patients aged 43.9 (18.8–66.81: range) years and 19 healthy subjects aged 34.9 (18.2–59.7: range) years with matching competitive levels were included (Table 1). There was a 16 ± 15 day difference between both testing moments. Age, length and approximate lifetime cycling kilometres were not different between groups. However, patients had a significantly higher BMI and attained lower peak workloads when normalized for body weight (W_{\max}/kg) compared to healthy controls. Absolute and relative reliability for W_{\max}/kg

TABLE 1 Characteristics of healthy subjects and patients

	Healthy (n = 19)	Patients (n = 17)
Gender (males)	17 (89%)	13 (76%)
Age (years) ^a	34.9 (18.2–59.7)	43.9 (18.8–66.8)
Length (cm)	179.2 (8.2)	179.9 (8.9)
Weight (kg)	71.1 (7.6)	81.6 (15.3) ^b
BMI (kg/m ²)	22.1 (2.1)	25.0 (2.9) ^b
Kilometre cycled in their life	206 000 (151 000)	260 000 (201 000)
Sports level		
Professional	2 (11%)	2 (12%)
Competitive	8 (42%)	6 (35%)
Recreational	9 (47%)	9 (53%)
Watt/kg	5.79 (0.94)	4.16 (0.89)
ABI _{flexed}	0.75 (0.12)	0.69 (0.18) ^b
Kilometer cycled in their life	206 000 (151 000)	260 000 (201 000)
PSV extended (m/s)	0.93 (0.21)	1.41 (0.48) ^b
PSV flexed (m/s)	1.12 (0.31)	1.87 (0.77) ^b
PSV psoas contraction (m/s)	1.32 (0.40)	2.36 (0.97) ^b

Note: Normal values for PSV extended, flexed and psoas contractions were ≤ 1.48 , ≤ 1.70 and ≤ 1.72 , respectively; normal value for ABI_{flexed} in competitive posture is >0.54 (Schip, Bender, Schmikli, et al., 2002; Schep, Bender, van de Tempel, et al., 2002; Schep, Schmikli, et al., 2002).

Abbreviations: ABI_{flexed}, ankle-brachial index with hips flexed; BMI, body mass index; PSV, peak systolic velocity.

^aAge is presented as the mean with its range.

^b $p < 0.05$.

were perfect. The modified ABI was normal in both groups suggesting subtle flow limitations in the patients. ABI showed good to excellent absolute reliability and substantial relative reliability. Peak systolic velocity (PSV) values were normal for patients while in supine posture. However, PSV became abnormal in a flexed hip position and even increased with active isometric iliopsoas contractions, indicating that any flow limitation is predominantly caused by functional kinking (Table 1). Despite a normal ABI_{flexed} in patients, a significant difference is found compared to the healthy subjects.

4.2 | NIRS measurements

During analysis, a small portion of the data had to be discarded for several reasons. In a few cases, data recording was involuntarily terminated before reaching a stable endpoint maximum, so recovery amplitude, HVT and monoexponential fit could not be determined (e.g., for the calculation of MRT of [dHb]), data of two patients were omitted). During recovery, two patients showed no reoxygenation for 5 min after ceasing exercising while maintaining competitive posture.

Both were diagnosed with complete unilateral occlusion of the iliac artery by MRA and Echo-Doppler. One of these patients was previously described in detail (van Hooff et al., 2020). Despite the implications for diagnostic purposes, no kinetics analysis was possible in these two cases and were, therefore, omitted from the analysis. A third patient with severe intravascular narrowing showed an extremely delayed reoxygenation pattern that did not follow an exponential increase preventing accurate fitting. The results of this patient were also omitted from the statistical analysis of reliability.

Using data from all three source-detector distances of the NIRS probe resulted in faulty calculated TSI due to inaccurate (HHb). Specifically, the largest source-detector distance often could not produce a reliable signal at all, which affected the interpretation across the three combined signals. This phenomenon may have been caused by muscle hypertrophy and excessive (or “elevated”) blood volume related to the highly trained status of our subjects. As such, this could have caused (near) complete absorption resulting in insufficient saturation of the photodetector

TABLE 2 Values of the TSI, [O₂Hb] and [dHb] of both days

	Parameter	Healthy			Patients			Sig. diff.
		N (legs)	Day 1	Day 2	N (legs)	Day 1	Day 2	
Cycling test	W_{\max} /kg	19	5.79 (0.94)	5.83 (1.06)	17	4.16 (0.75)	4.21 (0.8) ^a	b,c
	ABI _{flexed}	34	0.75 (0.12)	0.78 (0.1) ^a	29	0.69 (0.18)	0.70 (0.2)	b
NIRS absolute values and amplitudes								
TSI	Baseline (%)	36	62.2 (5.6)	61.7 (5.0)	30	61.3 (5.6)	60.3 (6.3)	
	Minimal (%)	32	43.5 (7.1)	42.4 (8.0)	27	44.2 (9.9)	42.5 (8.5)	
	Maximal (%)	32	67.8 (4.1)	67.2 (6.1)	22	68.9 (3.9)	68.2 (3.6)	
	Δ_{Exercise} (%)	32	18.5 (6.9)	19.1 (8.1)	27	17.5 (8.7)	18.5 (8.3)	
	Δ_{Recovery} (%)	28	24.4 (7.8)	25.5 (8.5)	21	24.9 (9.9)	25.4 (9.6)	
[O ₂ Hb]	Maximal (a. u.)	36	7.5 (6.4)	8.4 (6.0)	23	4.8 (5.7)	4.6 (6.1)	b
	Δ_{Exercise} (a. u.)	37	15.0 (4.7)	16.6 (5.9) ^a	30	14.3 (6.8)	15.6 (6.8)	
	Δ_{Recovery} (a. u.)	36	22.4 (8.0)	24.7 (8.0) ^a	23	18.1 (7.5)	19.2 (8.4)	b,c
[dHb]	Maximal (a. u.)	33	8.9 (7.5)	9.7 (7.0)	22	5.7 (6.7)	6.7 (6.7)	
	Δ_{Exercise} (a. u.)	32	24.1 (8.6)	26.6 (10.1)	29	24.1 (13.5)	24.9 (12.2)	
	Δ_{Recovery} (a. u.)	28	31.3 (11.9)	34.4 (12.9) ^a	22	29.2 (13.8)	32.4 (14.5)	
NIRS kinetic variables								
TSI	τ (s)	35	18.6 (12.6)	19.3 (9.4)	24	26.3 (11.4)	24.0 (11.4)	b
	Td (s)	35	8.7 (8.2)	9.7 (7.0)	24	15.8 (10.1)	15.1 (10.2)	b,c
	MRT (s)	35	27.3 (17)	28.9 (13.5)	24	42.1 (19.5)	39.1 (17.6)	b,c
	HVT (s)	27	21.4 (12.9)	22.6 (7.4)	21	32.4 (14.0)	29.0 (10.7)	b
[O ₂ Hb]	τ (s)	36	22.6 (11.5)	24.1 (10.2)	25	37.1 (24.7)	33.3 (15.9)	b,c
	Td (s)	36	9.1 (8.8)	10.3 (6.8)	25	20.4 (13.6)	20.5 (17.7)	b,c
	MRT (s)	36	31.7 (14.8)	34.3 (11.9)	25	57.5 (32.4)	53.7 (27.1)	b,c
	HVT (s)	35	24.4 (10.5)	24.9 (7.6)	22	40.2 (14.5)	36.7 (12.5)	b,c
[dHb]	τ (s)	34	24.9 (18.6)	27.1 (16.6)	25	35.3 (15.2)	34.2 (15.5)	b,c
	Td (s)	34	10.1 (7.5)	11.0 (8.6)	25	21.6 (17.1)	16.2 (8.3)	b,c
	MRT (s)	34	35.0 (22.4)	38.2 (22.5)	25	56.9 (28.0)	50.4 (20.7)	b,c
	HVT (s)	27	23.4 (8.5)	25.8 (7.7)	21	38.3 (13.2)	37.0 (12.1)	b,c

Note: Data are presented for both legs of healthy subjects and patients. Values are mean \pm standard deviation. Kinetic values are expressed in seconds. Abbreviations: ABI_{flexed}, ankle-brachial index with hips flexed; MRT, mean response time; τ , is the time constant of the monoexponential model; W_{\max} /kg, maximal attained workload per kilogram body weight; Td, time delay.

^aBetween test days significant difference.

^bSignificant difference between healthy subjects and patients on Day 1.

^cSignificant difference between healthy subjects and patients on Day 2.

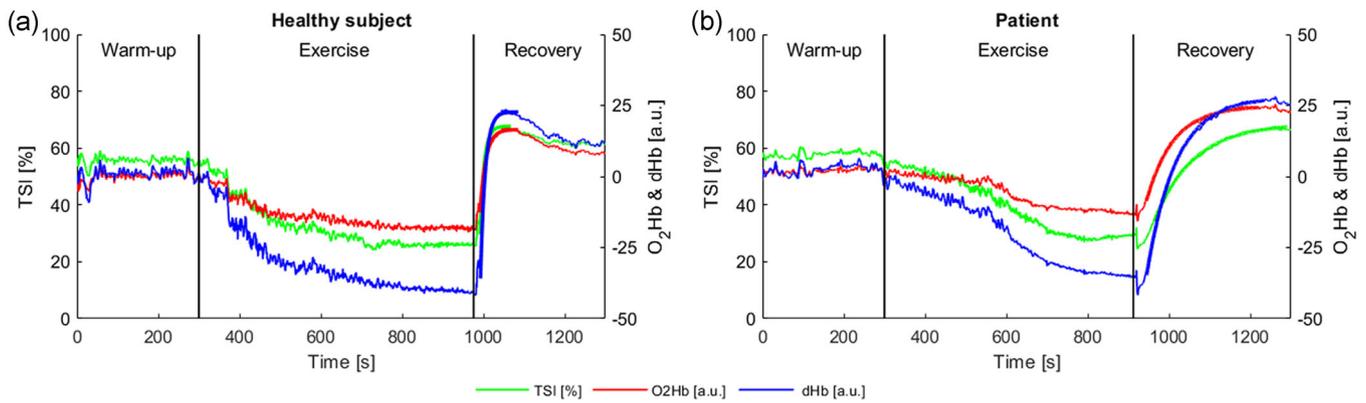


FIGURE 3 Visual presentation of the output of NIRS signals includes TSI, [O₂Hb] and [dHb] in (a) a healthy subject and (b) a patient with functional arterial kinking. The thick solid in the recovery phase represents the monoexponential model. NIRS, near-infrared spectroscopy; TSI, Tissue Saturation Index

with scattered light. For this reason, only the signals from two LEDs (30–35 source-detector distance) were used for all datasets, as per manufacturer recommendations. All fits analyzed were considered satisfactory ($R^2 > 0.85$).

4.3 | Healthy versus patients with FLIA

There was a nearly identical maximal workload achieved by all subjects during both testing days. The body-mass normalized maximal workload of patients was significantly lower compared to healthy subjects (Table 2). A significantly lower ABI and higher difference of pressures between the ankles of patients compared to healthy subjects was found. Neither the absolute values nor amplitudes showed significant differences, except for the maximal value and Δ_{recovery} in [O₂Hb], between patients and healthy subjects. In contrast, significant differences were found between patients and healthy subjects regarding almost all reoxygenation kinetics (e.g., [O₂Hb] - MRT in healthy subjects of 31.7 ± 14.8 s vs. 57.5 ± 32.4 s in patients). In Figure 2, the different kinetic parameters (τ , T_d , and MRT) are visualized, whereas a clear difference is seen between the healthy subject (Figure 3A) and the patient (Figure 3B). Significant but moderate correlations between ABI and most kinetic parameters were found (delay, MRT and HVT range -0.27 to -0.42 , τ -0.04 to -0.15). However, no relation was found between ABI and absolute values (0.00 – 0.22) with the exception of from O₂Hb Δ_{Exercise} ($r = -0.35$) This might be attributed to the relatively subtle flow-limitations in combination with a normal ABI which is often the case in these patient populations. As earlier described, Schep et al. found that in the rare case FLIA is unilateral, the sensitivity of this test is 73%. When a bilateral FLIA is present, sensitivity is only 43% (Schep, Bender, Schmikli, et al., 2002; Schep, Bender, van de Tempel, et al., 2002; Schep, Schmikli, et al., 2002) (Table 3).

4.4 | Test-rest reliability

Visual inspection of Bland–Altman plots (Figure 4) and their logarithmic transformations revealed heteroskedasticity in several variables, making the RLoA more representative. Regarding baseline, minimal and maximal values, the test–retest reliability showed slight to substantial relative reliability (ICC_{Healthy} 0.16–0.70; ICC_{FLIA} 0.06–0.76; ICC_{Total} 0.14–0.73) with good to excellent absolute reliability (ALoA_{healthy} -0.9 ± 10.5 to 0.7 ± 13.1 ; CV_{healthy}; 6%–10%; ALoA_{FLIA} -0.8 ± 10.2 to 1.1 ± 12.8 ; CV_{FLIA} 5%–11%). Substantial to almost perfect relative reliability (ICC_{Healthy} 0.51–0.83; ICC_{FLIA} 0.80–0.89; ICC_{Total} 0.66–0.84) was found in signal amplitudes (Δ_{exercise} and Δ_{recovery}) and poor to good absolute reliability (ALoA_{healthy} -2.3 ± 8.4 to -2.5 ± 15.7 ; CV_{healthy}; 15%–28%; ALoA_{FLIA} -1.3 ± 7.0 to 3.2 ± 14.6 ; CV_{FLIA} 16%–32%). Regarding the recovery kinetic variables, we found a moderate to almost perfect agreement (ICC_{Healthy} 0.54–0.86; ICC_{FLIA} 0.43–0.81; ICC_{Total} 0.56–0.84). However, the ALoA showed relatively large limits and fair to good CV (ALoA_{healthy} 0.4 ± 12.2 to 3.2 ± 22.7 ; CV_{healthy}; 19%–27%; ALoA_{FLIA} 6.4 ± 15.7 to 3.9 ± 37.9 ; CV_{FLIA} 18%–27%). Interestingly, a few variables ([O₂Hb] Δ_{exercise} , [O₂Hb] Δ_{recovery} , and [dHb] Δ_{recovery}) variables showed significant differences between the testing days (Table 2).

5 | DISCUSSION

The ABI (ABI_{flexed}) is currently the best available functional variable to diagnose FLIA in athletes, but this measurement lacks diagnostic accuracy (Abraham et al., 2001; Schep, Bender, Schmikli, et al., 2002; Schep, Bender, van de Tempel, et al., 2002; Schep, Schmikli, et al., 2002). The aim of this study was to assess the test–retest reliability of possible relevant NIRS variables and to explore if this technique has the potential to improve the diagnosis of FLIA during and after an incremental maximal exercise test. During this exercise and recovery, several absolute variables and

TABLE 3 Test-retest reliability of TSI, [O₂Hb] and [dHb] for both testing days

Variable	Healthy				Patient				Total group		
	N	ALoA	Ratio	CV	ICC (95% CI)	N	ALoA	Ratio	CV	ICC (95% CI)	ICC (95% CI)
Cycling test	19	-0.05 (0.66)		4%	0.95 (0.87-0.98)	17	-0.06 (0.33)		4%	0.98 (0.94-0.99)	0.98 (0.95-0.99)
ABI	34	-0.03 (0.17)		9%	0.67 (0.42-0.82)	29	-0.01 (0.3)		26%	0.68 (0.42-0.84)	0.69 (0.54-0.80)
Absolute values and signal amplitudes											
TSI	36	0.5 (11.2)		6%	0.42 (0.11-0.66)	30	1 (11.4)		7%	0.52 (0.21-0.74)	0.47 (0.27-0.64)
Minimal (%)	32	1.1 (11.6)		10%	0.70 (0.46-0.84)	27	1.7 (12.5)		11%	0.76 (0.54-0.88)	0.73 (0.58-0.83)
Maximal (%)	32	0.7 (13.1)		8%	0.16 (-0.21 to 0.48)	22	0.8 (10.2)		5%	0.06 (-0.37 to 0.47)	0.14 (-0.13 to 0.39)
Δ _{Exercise} (%)	32	-0.7 (14.7)	0.99x/% 2.40	28%	0.51 (0.20-0.73)	27	-1 (10.5)		32%	0.80 (0.62-0.91)	0.66 (0.49-0.78)
Δ _{Recovery} (%)	28	-1.1 (14.9)	0.96x/% 1.92	22%	0.57 (0.26-0.77)	21	-0.5 (9.3)		16%	0.89 (0.74-0.95)	0.73 (0.57-0.84)
[O ₂ Hb]	36	-0.9 (10.5)	^a	^a	0.63 (0.39-0.79)	23	0.2 (10.8)	^a	^a	0.57 (0.21-0.79)	0.63 (0.45-0.76)
Δ _{Exercise}	37	-1.6 (8.9)	0.92x/% 1.80	21%	0.62 (0.38-0.79)	30	-1.3 (7.0)		23%	0.85 (0.70-0.93)	0.75 (0.60-0.84)
Δ _{Recovery}	36	-2.3 (8.4)	0.90x/% 1.47	16%	0.83 (0.62-0.92)	23	-1.1 (9.4)	0.96x/% 1.71	19%	0.82 (0.62-0.92)	0.83 (0.71-0.90)
[dHb]	33	-0.8 (11.7)	^a	^a	0.67 (0.42-0.82)	22	-1.1 (12.8)	^a	^a	0.52 (0.14-0.77)	0.63 (0.44-0.76)
Δ _{Exercise}	32	-2.5 (15.7)	0.92x/% 1.79	21%	0.62 (0.36-0.80)	29	-0.8 (13.0)	0.95x/% 1.68	18%	0.87 (0.74-0.94)	0.77 (0.65-0.86)
Δ _{Recovery}	28	-3.1 (13.2)	0.91x/% 1.47	15%	0.83 (0.64-0.92)	22	-3.2 (14.6)	0.90x/% 1.66	19%	0.85 (0.65-0.93)	0.84 (0.69-0.91)
Kinetic variables											
TSI	35	-0.6 (18.8)	0.93x/% 2.14	27%	0.63 (0.38-0.80)	24	2.3 (23.7)	1.12x/% 2.11	27%	0.44 (0.06-0.71)	0.58 (0.38-0.73)
Td (s)	35	-1.0 (14.4)		^a	0.54 (0.25-0.74)	24	0.7 (15.7)		^a	0.70 (0.42-0.86)	0.66 (0.49-0.79)
MRT (s)	35	-1.6 (22.6)	0.90x/% 2.04	25%	0.72 (0.51-0.85)	24	3.0 (27.6)	1.06x/% 1.78	20%	0.71 (0.44-0.86)	0.75 (0.61-0.84)
HVT (s)	27	-1.2 (18.7)	0.89x/% 2.06	26%	0.59 (0.28-0.79)	20	3.4 (20.3)	1.09x/% 1.84	21%	0.64 (0.30-0.84)	0.66 (0.46-0.79)
[O ₂ Hb]	36	-1.5 (16.6)	0.93x/% 2.01	25%	0.70 (0.48-0.83)	25	3.9 (36.6)	1.08x/% 2.06	26%	0.60 (0.27-0.80)	0.66 (0.50-0.78)
Td (s)	36	-1.2 (12.8)		^a	0.65 (0.42-0.81)	25	0.0 (25.4)	1.08x/% 2.36	^a	0.67 (0.38-0.84)	0.72 (0.57-0.82)
MRT (s)	36	-2.7 (17.7)	0.89x/% 1.71	20%	0.76 (0.58-0.87)	25	3.8 (36.2)	1.07x/% 1.66	18%	0.81 (0.62-0.91)	0.84 (0.74-0.90)
HVT (s)	35	-0.4 (12.2)	0.95x/% 1.73	19%	0.77 (0.60-0.88)	22	3.5 (21.5)	1.09x/% 1.67	19%	0.66 (0.35-0.84)	0.78 (0.66-0.87)
[dHb]	34	-2.2 (18.2)	0.89x/% 1.69	20%	0.86 (0.74-0.93)	25	1.1 (23.4)	1.05x/% 1.75	20%	0.70 (0.43-0.86)	0.81 (0.70-0.88)
Td (s)	34	-1.0 (12.7)		^a	0.68 (0.45-0.83)	25	5.4 (27.6)		^a	0.43 (0.07-0.69)	0.56 (0.36-0.71)
MRT (s)	34	-3.2 (22.7)	0.90x/% 1.65	19%	0.86 (0.74-0.93)	25	6.4 (37.9)	1.10x/% 1.75	21%	0.68 (0.40-0.84)	0.80 (0.68-0.87)
HVT (s)	27	-2.3 (13.1)		22%	0.64 (0.36-0.82)	21	1.4 (20.8)	1.03x/% 1.67	18%	0.66 (0.32-0.85)	0.75 (0.59-0.85)

Abbreviations: ALoA, absolute limit of agreement expressed as the mean difference ± random error; ABI_{flexed}, ankle-brachial index with hips flexed; CV, coefficient of variation; ICC, intraclass correlation coefficient; MRT, mean response time; Tau is the time constant of the monoexponential model; Td is the time delay; W_{max}/kg, maximal attained workload per kilogram body weight.

^aIn case of heteroscedasticity, the logarithmically transformed ratio values are shown (in cases of negative and positive values prohibit logarithmical transformation and meaningful CV).

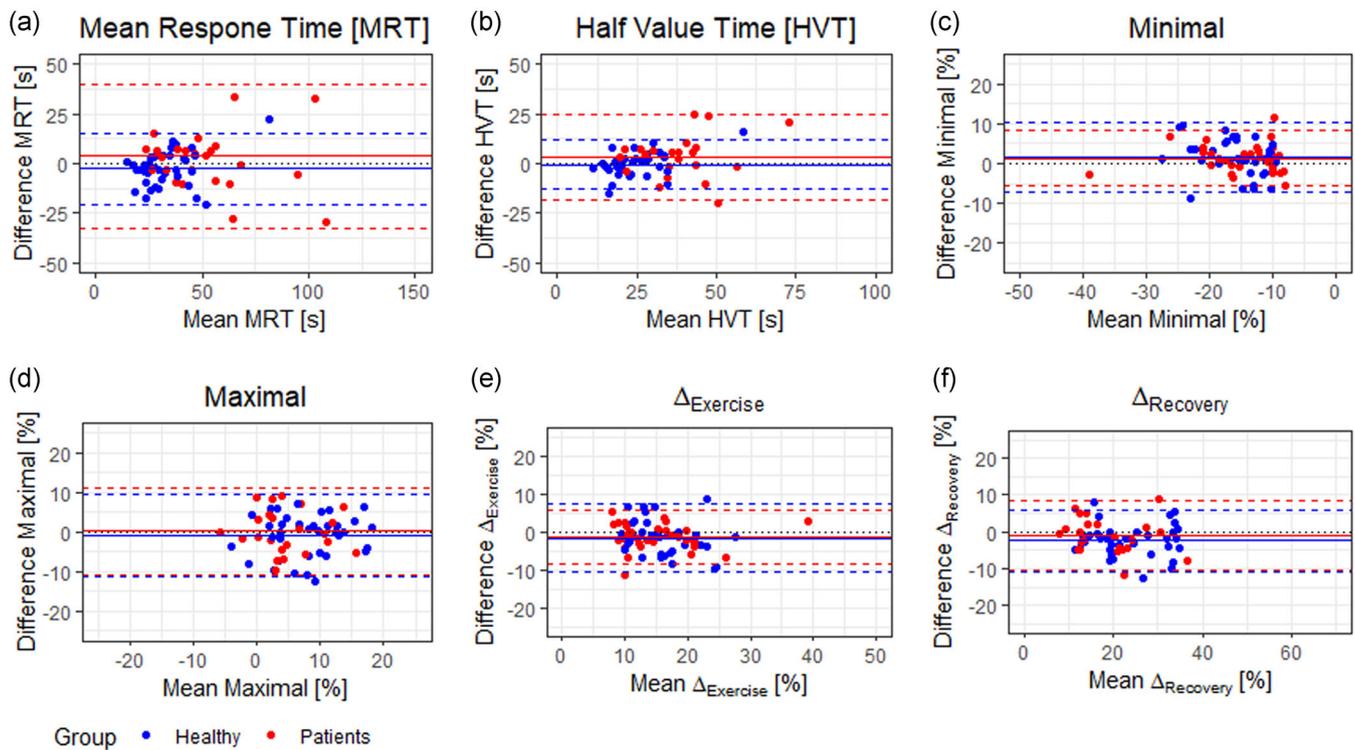


FIGURE 4 Bland–Altman showing the difference between the incremental maximal exercise test of both moments against the average of the two tests for TSI. The solid line represents the mean difference between the two tests while the dashed lines indicate the 95 confidence interval of the differences. TSI, Tissue Saturation Index

signal amplitudes were examined and reoxygenation kinetics were modelled.

We found that the absolute reliability (the similarity of repeated measures) and relative reliability (the ability to assess differences between subjects while taking the measurement error into account) were good to almost perfect in most absolute NIRS variables and amplitudes, but with only minor significant differences between patients and healthy subjects. By contrast, most recovery kinetic (τ , MRT and HVT) variables showed fair to good absolute reliability yet good to almost perfect relative reliability with statistically significant differences in the responses of recovery kinetics after exercise between FLIA patients having predominantly subtle flow limitations and healthy subjects. Other studies showed that absolute values, amplitude and reoxygenation kinetics showed poor to near excellent relative reliability in varying populations and exercise types (Baláš et al., 2018; Buchheit et al., 2011; Cayot et al., 2021; Choo et al., 2016; Contreras-Briceño et al., 2019; Crenshaw et al., 2012; Ihsan et al., 2013; Leclair et al., 2010; Lucero et al., 2018; McManus et al., 2018; Miranda-Fuentes et al., 2020; Muthalib et al., 2010; Niemeijer et al., 2017; Pocivalnik et al., 2011; Thiel et al., 2011; Ubbink & Koopman, 2006). Given the combination of good to almost perfect relative reliability and significant differences in reoxygenation variables, results suggest that NIRS recovery kinetics (MRT and HVT) may be useful and may prove superior to ABl_{flexed} when diagnosing sport-related flow limitations in the iliac arteries. As the

absolute reliability was near the lower bounds of acceptability, the potential of using recovery kinetics for monitoring a subject over time may be limited.

5.1 | Exercise testing

There was a significant difference in the maximal workload when normalized to body weight (W/kg) between FLIA patients and healthy subjects. This is most likely due to the flow limitation in the patients and/or the (nonsignificant) age difference leading to a decrease of physical fitness. A difference in fitness may also explain the significantly higher weight (and BMI) in patients.

5.2 | Baseline

Several studies have assessed NIRS variables during exercise with varying test–retest reliability. When considering the TSI at rest in the VL muscle, others found values of CV to range from 1% to 6% and ICC from 0.71 and 0.92 (Cayot et al., 2021; Choo et al., 2016; Ihsan et al., 2013; Lucero et al., 2018; McManus et al., 2018; Niemeijer et al., 2017; Thiel et al., 2011). Our study finds a slightly higher CV (6% and 7%) and a lower ICC_{Total} (0.47). The limits of agreement were comparable with Thiel et al., but higher than Niemeijer et al. (Niemeijer et al., 2017; Thiel et al., 2011). These differences are likely

due to different measurement protocols and body positions. Our resting period for baseline assessment consisted of 1 min seated on the bike without restrictions on pedal position. Other studies used different techniques or muscles (e.g., multiple sensors, upper extremities) with a large range of ICC (0.42–0.95) and CV (2%–8%) (Baláš et al., 2018; Contreras-Briceño et al., 2019; Crenshaw et al., 2012; Miranda-Fuentes et al., 2020; Muthalib et al., 2010; Ubbink & Koopman, 2006).

5.3 | Absolute values

The minimum TSI at maximal exercise in the VL showed higher ICC (0.87–0.93) and comparable CV (11%) to our study group (ICC_{Healthy} of 0.70, ICC_{FLIA} of 0.76 and 0.73 for ICC_{Total}; CV of 10% and 11%) (Cayot et al., 2021; Thiel et al., 2011). Our exercise signal amplitude showed fair to predominantly good absolute reliability with substantial to almost perfect relative reliability (CV range: 13%–28%; ICC_{Total} range: 0.66–0.84). Therefore, these absolute values and amplitudes may be valuable for follow-up studies. However, caution should be taken with Δ_{Exercise} in $[\text{O}_2\text{Hb}]$ and Δ_{Recovery} in both $[\text{O}_2\text{Hb}]$ and $[\text{dHb}]$ as they revealed significant differences between testing days which questions the usability. Niemeijer et al. found inferior CV values yet better ICC values for TSI in patients with chronic heart failure (CHF) for Δ_{Exercise} (CV 42% vs. 28% and 32%; ICC 0.85 vs. 0.66 for ICC_{Total}) and Δ_{Recovery} (CV 26% vs. 22% and 16%; ICC 0.92 vs. 0.73 for ICC_{Total}) (Niemeijer et al., 2017). The present results suggest that these variables are able to assess differences between subjects and may have the potential to monitor a patient. However, given that there were minor differences between healthy subjects and patients, the diagnostic applicability for FLIA seems limited.

5.4 | Kinetics analysis during recovery

Since patients with FLIA are likely to have slower reoxygenation during recovery, kinetics are of great interest. In one study comparing patients with CHF with healthy subjects, recovery kinetics in the VL and medial gastrocnemius muscles following sprint exercise showed quite poor relative reliability (ICC for TSI of 0.51, $[\text{O}_2\text{Hb}]$ of 0.50 and $[\text{dHb}]$ of 0.30) for the MRT (Buchheit et al., 2011; Niemeijer et al., 2017). Our study, however, found superior results for ICC of TSI, $[\text{O}_2\text{Hb}]$ and $[\text{dHb}]$ (0.72, 0.76 and 0.86 for healthy subjects; 0.71, 0.81 and 0.68 for FLIA patients). In contrast, the absolute reliability was lower in our study compared to the patients with CHF regarding TSI (67% vs. 25% and 20% in our healthy subjects and patients) yet lower but comparable with the study in healthy subjects after a maximal sprint test ($[\text{O}_2\text{Hb}]$ 27% vs. 20% and 18% in our study and for $[\text{dHb}]$ 33% vs. 19% and 21%) (Buchheit et al., 2011; Niemeijer et al., 2017).

Interestingly, the time delay and τ had lower relative reliability and broader limits of agreement compared to the summation of the two as MRT. This phenomenon is also seen in one other

study (Leclair et al., 2010). This may be the result of ambiguous fitting parameters around the initial inflection point of the mono-exponential function, related to physiological variability and/or movement disruption during the brief transition from pedal to platform. Yet, the majority of the evaluated kinetics variables show substantial to almost perfect relative reproducibility. This finding suggests that these measures are able to assess differences between subjects. However, giving the broader (ratio) limits of agreement, the use of these outcome measures for longitudinal patient monitoring may be limited.

6 | STUDY LIMITATIONS

First, this study included only four women in the patient group and two in the control group. This limits the generalization of the results to females. In general, females tend to have higher ATT at the quadriceps measurement site, which may also limit the quality of NIRS signals that are collected in female subjects. Second, only one type of NIRS device was used. Some research has reported differences between devices, even when provided by one manufacturer. Therefore, caution should be taken when comparing reported values between studies, especially using different NIRS sensors and technologies (Hyttel-Sorensen et al., 2011; Pocivalnik et al., 2011). Third, only one muscle region was investigated. Multiple site analysis may reveal local variations in perfusion and oxygen consumption and could, therefore, improve diagnostic sensitivity and test–retest reliability (Koga et al., 2007). The development of multiopode measurement across a larger three-dimensional tissue volume could lead to more robust measurements in the future. Fourth, only two tests were performed for analysis; whereas multiple subsequent tests might give more satisfactory results for individual patients (absolute agreement) as shown by Spencer et al. (2011). Fifth, in the absence of a validated pathophysiological model of reoxygenation, we chose a monoexponential function that has been used extensively in the literature providing a robust fit to our data in most cases. Finally, placement and pressure on the NIRS device could have influenced measurements. The pencil markings made during the first trial were not present at the second session and differences in the attachment of the fabric strap holding the NIRS sensor may have changed the pressure of the device on the skin. It is known that compression influences oxygen saturation and availability (Bringard et al., 2006; Coza et al., 2012). Nonetheless, caution was taken to limit these factors.

7 | TECHNICAL LIMITATIONS

There were some technical limitations related to our population characteristics. Our subjects and patients were generally well trained with low ATT levels and hypertrophy of the lower extremity muscles. Unfortunately, this in combination with the elevated blood volume resulted in complete absorption of the NIRS optical signal, making

especially TSI and [dHb] unreliable for calculations. However, the manufacturer expects that this drawback will be mitigated with a new feature called multi power gain control, which automatically enhances the exposure time of the LEDs that received insufficient light to make reliable calculations.

8 | CLINICAL APPLICABILITY AND IMPLICATIONS FOR FUTURE RESEARCH

NIRS may be a valuable tool for the clinical assessment of patients with FLIA. It may prove useful in the process of diagnosis, classification of impairment severity, and for monitoring of outcome measures, for example, for the evaluation of conservative interventions (e.g., posture adaptation) and surgical treatment. Studies using NIRS have shown delayed reoxygenation kinetics and deoxygenation rate in patients with the peripheral arterial disease (PAD) (Cheatle et al., 1991; Kemp et al., 2001; Komiyama et al., 2000). However, limitations of NIRS for diagnosing patients with PAD are population related (e.g., obesity, cellulitis, oedema), overlapping variability between healthy subjects and patients, and a lack of a standardized approach (Baltrunas et al., 2021). A future larger study is necessary to more precisely delineate the diagnostic properties (diagnostic thresholds, sensitivity and specificity) of NIRS recovery variables in patients with FLIA, and its additional diagnostic properties compared to currently used methods. However, current differences in exercise mode (treadmill or bike), intensity (continuous workload or incremental exercise), NIRS device type (spatially resolved spectroscopy, continuous wave, or frequency domain) and position of the device hamper comparison among studies. This heterogeneity of methods in the literature highlights the need for consistent practices when it comes to establishing a standard clinical diagnostic approach and outcome monitoring criteria for FLIA.

Our results for NIRS during exercise may be more improved with simultaneous measures of gas exchange and determination of individual intensity domains, which could give more insight into the cause and mechanisms for variation of the NIRS signals. As NIRS devices are constantly being developed and improved and more commercially available devices are introduced in the field, important physiological information for athletes and coaches will become available.

9 | CONCLUSION

This is the first research investigating the absolute and relative test–retest reliability of variables obtained with NIRS during and after an incremental ramp maximal cycling exercise test in healthy subjects and patients with sport-related flow limitations. Absolute values and amplitudes reveal both good absolute and relative reliability. However, limited significant differences were found between the groups. Recovery kinetics shows that absolute test–retest reliability reveals

broad (ratio) limits of agreement. The potential use of NIRS to monitor a person over a period of time, or testing differences in posture during cycling may therefore be limited. However, the kinetic variables show substantial to almost perfect relative test–retest reliability with the statistical difference between healthy subjects and patients. Therefore, this technique is useful for screening and diagnostic purposes of patients having FLIA. Further research is necessary to determine diagnostic thresholds/criteria and diagnostic accuracy.

ACKNOWLEDGEMENTS

We would like to thank Nina van de Rijdt, Paul Schreuder, Bart Smeets, Lianne Houben and Noortje Tabak with the help of subject inclusion, data acquisition and data analysis. Moreover, we would like to thank Victor Niemeijer for his effort in reading the manuscript given his expertise in test–retest reliability using NIRS in patients with chronic heart failure. This study was partly supported by ZonMw [Project no. 75020017].

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Martijn van Hooff  <http://orcid.org/0000-0002-8976-3500>

Eduard J. Meijer  <http://orcid.org/0000-0002-4753-8671>

Marc R. M. Scheltinga  <http://orcid.org/0000-0003-4924-6194>

Hans H. C. M. Savelberg  <http://orcid.org/0000-0002-8864-2109>

Goof Schep  <http://orcid.org/0000-0003-0207-8066>

REFERENCES

- Abraham, P., Bickert, S., Vielle, B., Chevalier, J.M. & Saumet, J.L. (2001) Pressure measurements at rest and after heavy exercise to detect moderate arterial lesions in athletes. *Journal of Vascular Surgery*, 33, 721–727.
- Abraham, P., Chevalier, J.M., Leftheriotis, G. & Saumet, J.L. (1997) Lower extremity arterial disease in sports. *American Journal of Sports Medicine*, 25, 581–584.
- Atkinson, G. & Nevill, A.M. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26, 217–238.
- Baláš, J., Kodejška, J., Krupková, D., Hannsmann, J. & Fryer, S. (2018) Reliability of near-infrared spectroscopy for measuring intermittent handgrip contractions in sport climbers. *The Journal of Strength and Conditioning Research*, 32, 494–501.
- Baltrunas, T., Mosenko, V., Mackevicius, A., Dambrauskas, V., Asakiene, I., Rucinskas, K. et al. (2021) The use of near-infrared spectroscopy in the diagnosis of peripheral artery disease: a systematic review. *Vascular*, 1–13. <https://doi-org.ezproxy.ub.unimaas.nl/10.1177/17085381211025174>
- Bland, J.M. & Altman, D.G. (1999) Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135–160.

- Bland, J.M. & Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307–310.
- Bringard, A., Denis, R., Belluye, N. & Perrey, S. (2006) Effects of compression tights on calf muscle oxygenation and venous pooling during quiet resting in supine and standing positions. *The Journal of Sports Medicine and Physical Fitness*, 46, 548–554.
- Buchheit, M., Ufland, P., Haydar, B., Laursen, P.B. & Ahmadi, S. (2011) Reproducibility and sensitivity of muscle reoxygenation and oxygen uptake recovery kinetics following running exercise in the field. *Clinical Physiology and Functional Imaging*, 31, 337–346.
- Cayot, T.E., Robinson, S.G., Davis, L.E., Bender, P.A., Thistlethwaite, J.R., Broeder, C.E. et al. (2021) Estimating the lactate threshold using wireless near-infrared spectroscopy and threshold detection analyses. *International Journal of Exercise Science*, 14, 284–294.
- Cheatle, T.R., Potter, L.A., Cope, M., Delpy, D.T., Coleridge Smith, P.D. & Scurr, J.H. (1991) Near-infrared spectroscopy in peripheral vascular disease. *British Journal of Surgery*, 78, 405–408.
- Chevalier, J.M., Enon, B., Walder, J., Barral, X., Pillet, J., Megret, A. et al. (1986) Endofibrosis of the external iliac artery in bicycle racers: an unrecognized pathological state. *Annals of Vascular Surgery*, 1, 297–303.
- Choo, H.C., Nosaka, K., Peiffer, J.J., Ihsan, M., Yeo, C.C. & Abbiss, C.R. (2016) Reliability of laser Doppler, near-infrared spectroscopy and Doppler ultrasound for peripheral blood flow measurements during and after exercise in the heat. *Journal of Sports Sciences*, 35, 1–9.
- Cicchetti, D. (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. *Psychological Assessment*, 6, 284–290.
- Contreras-Briceño, F., Espinosa-Ramírez, M., Hevia, G., Llambias, D., Carrasco, M., Cerda, F. et al. (2019) Reliability of NIRS portable device for measuring intercostal muscles oxygenation during exercise. *Journal of Sports Sciences*, 37, 2653–2659.
- Coza, A., Dunn, J.F., Anderson, B. & Nigg, B.M. (2012) Effects of compression on muscle tissue oxygenation at the onset of exercise. *The Journal of Strength and Conditioning Research*, 26, 1631–1637.
- Crenshaw, A.G., Elcadi, G.H., Hellstrom, F. & Mathiassen, S.E. (2012) Reliability of near-infrared spectroscopy for measuring forearm and shoulder oxygenation in healthy males and females. *European Journal of Applied Physiology*, 112, 2703–2715.
- Cui, W., Kumar, C. & Chance, B. (1991) Experimental study of migration depth for the photons measured at sample surface. *SPIE*. 1431, 180–191.
- de Groote, P., Millaire, A., Decoulx, E., Nugue, O. & Guimier, P. (1996) Kinetics of oxygen consumption during and after exercise in patients with dilated cardiomyopathy. New markers of exercise intolerance with clinical implications. *Journal of the American College of Cardiology*, 28, 168–175.
- de Vet, H.C., Terwee, C.B., Knol, D.L. & Bouter, L.M. (2006) When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, 59, 1033–1039.
- Ferrari, M., Muthalib, M. & Quaresima, V. (2011) The use of near-infrared spectroscopy in understanding skeletal muscle physiology: recent developments. *Philosophical Transactions of the Royal Society A*, 369, 4577–4590.
- Ferreira, L.F., Harper, A.J., Townsend, D.K., Lutjemeier, B.J. & Barstow, T.J. (2005) Kinetics of estimated human muscle capillary blood flow during recovery from exercise. *Experimental Physiology*, 90, 715–726.
- Ferreira, L.F., Hueber, D.M. & Barstow, T.J. (2007) Effects of assuming constant optical scattering on measurements of muscle oxygenation by near-infrared spectroscopy during exercise. *Journal of Applied Physiology* (1985), 102, 358–367.
- Gornik, H.L., Garcia, B., Wolski, K., Jones, D.C., Macdonald, K.A. & Fronck, A. (2008) Validation of a method for determination of the ankle-brachial index in the seated position. *Journal of Vascular Surgery*, 48, 1204–1210.
- Grassi, B., Quaresima, V., Marconi, C., Ferrari, M. & Cerretelli, P. (1999) Blood lactate accumulation and muscle deoxygenation during incremental exercise. *Journal of Applied Physiology* (1985), 87, 348–355.
- Hyttel-Sorensen, S., Sorensen, L.C., Riera, J. & Greisen, G. (2011) Tissue oximetry: a comparison of mean values of regional tissue saturation, reproducibility and dynamic range of four NIRS-instruments on the human forearm. *Biomedical Optics Express*, 2, 3047–3057.
- Ihsan, M., Abbiss, C., Lipski, M., Buchheit, M. & Watson, G. (2013) Muscle oxygenation and blood volume reliability during continuous and intermittent running. *International Journal of Sports Medicine*, 34, 637–645.
- INSITE Collaborators (INternational Study group for Identification and Treatment of Endofibrosis). (2016) Diagnosis and management of iliac artery endofibrosis: results of a Delphi consensus study. *European Journal of Vascular and Endovascular Surgery*, 52, 90–98.
- Jones, R.G.M. & Payne, R.B. (1997). *Clinical investigation and statistics in laboratory medicine*. London, UK: ACB Venture Publications.
- Kemp, G.J., Roberts, N., Bimson, W.E., Bakran, A., Harris, P.L., Gilling-Smith, G.L. et al. (2001) Mitochondrial function and oxygen supply in normal and in chronically ischemic muscle: a combined ³¹P magnetic resonance spectroscopy and near infrared spectroscopy study in vivo. *Journal of Vascular Surgery*, 34, 1103–1110.
- Kemps, H.M., De Vries, W.R., Hoogeveen, A.R., Zonderland, M.L., Thijssen, E.J. & Schep, G. (2007) Reproducibility of onset and recovery oxygen uptake kinetics in moderately impaired patients with chronic heart failure. *European Journal of Applied Physiology*, 100, 45–52.
- Khan, A., Al-Dawoud, M., Salaman, R. & Al-Khaffaf, H. (2018) Management of endurance athletes with flow limitation in the iliac arteries: a case series. *EJVES Short Reports*, 40, 7–11.
- Koga, S., Poole, D.C., Ferreira, L.F., Whipp, B.J., Kondo, N., Saitoh, T. et al. (2007) Spatial heterogeneity of quadriceps muscle deoxygenation kinetics during cycle exercise. *Journal of Applied Physiology* (1985), 103, 2049–2056.
- Komiyama, T., Shigematsu, H., Yasuhara, H. & Muto, T. (2000) Near-infrared spectroscopy grades the severity of intermittent claudication in diabetics more accurately than ankle pressure measurement. *British Journal of Surgery*, 87, 459–466.
- Leclair, E., Thevenet, D., Reguem, S.C., Borel, B., Baquet, G., Berthoin, S. et al. (2010) Reproducibility of measurement of muscle deoxygenation in children during exercise. *Pediatric Exercise Science*, 22, 183–194.
- Lucero, A.A., Addae, G., Lawrence, W., Neway, B., Credeur, D.P., Faulkner, J. et al. (2018) Reliability of muscle blood flow and oxygen consumption response from exercise using near-infrared spectroscopy. *Experimental Physiology*, 103, 90–100.
- McManus, C.J., Collison, J. & Cooper, C.E. (2018) Performance comparison of the MOXY and PortaMon near-infrared spectroscopy muscle oximeters at rest and during exercise. *Journal of Biomedical Optics*, 23, 015007.
- Miranda-Fuentes, C., Guisado-Requena, I.M., Delgado-Floody, P., Arias-Poblete, L., Perez-Castilla, A. & Jerez-Mayorga, D. et al. (2020) Reliability of low-cost near-infrared spectroscopy in the determination of muscular oxygen saturation and hemoglobin concentration during rest, isometric and dynamic strength activity. *International Journal of Environmental Research and Public Health*, 17, 8824.
- Muthalib, M., Millet, G.Y., Quaresima, V. & Nosaka, K. (2010) Reliability of near-infrared spectroscopy for measuring biceps brachii oxygenation during sustained and repeated isometric contractions. *Journal of Biomedical Optics*, 15, 017008.
- Niemeijer, V.M., Jansen, J.P., van Dijk, T., Spee, R.F., Meijer, E.J., Kemps, H.M. et al. (2017) The influence of adipose tissue on spatially resolved near-infrared spectroscopy derived skeletal muscle

- oxygenation: the extent of the problem. *Physiological Measurement*, 38, 539–554.
- Niemeijer, V.M., Spee, R.F., Jansen, J.P., Buskermolen, A.B., van Dijk, T., Wijn, P.F. et al. (2017) Test-retest reliability of skeletal muscle oxygenation measurements during submaximal cycling exercise in patients with chronic heart failure. *Clinical Physiology and Functional Imaging*, 37, 68–78.
- O'Connor, F.G. (2013) *ACSM's sports medicine a comprehensive review*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Patterson, M.S., Chance, B. & Wilson, B.C. (1989) Time resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties. *Applied Optics*, 28, 2331–2336.
- Peach, G., Schep, G., Palfreeman, R., Beard, J.D., Thompson, M.M. & Hinchliffe, R.J. (2012) Endofibrosis and kinking of the iliac arteries in athletes: a systematic review. *European Journal of Vascular and Endovascular Surgery*, 43, 208–217.
- Pearson, R. (2002) Outliers in process modeling and identification. *Control Systems Technology, IEEE Transactions on*, 10, 55–63.
- Pocivalnik, M., Pichler, G., Zotter, H., Tax, N., Muller, W. & Urlesberger, B. (2011) Regional tissue oxygen saturation: comparability and reproducibility of different devices. *Journal of Biomedical Optics*, 16, 057004.
- Rickham, P.P. (1964) Human experimentation. Code of Ethics of the World Medical Association. Declaration of Helsinki. *British Medical Journal*, 2, 177.
- Rosner, B. (2016) *Fundamentals of biostatistics*. Boston, US: Cengage Learning.
- Schep, G., Bender, M.H., Schmikli, S.L., Mosterd, W.L., Hammacher, E.R., Scheltinga, M. et al. (2002) Recognising vascular causes of leg complaints in endurance athletes. Part 2: the value of patient history, physical examination, cycling exercise test and echo-Doppler examination. *International Journal of Sports Medicine*, 23, 322–328.
- Schep, G., Bender, M.H., Schmikli, S.L. & Wijn, P.F. (2001) Color Doppler used to detect kinking and intravascular lesions in the iliac arteries in endurance athletes with claudication. *European Journal of Ultrasound*, 14, 129–140.
- Schep, G., Bender, M.H., van de Tempel, G., Wijn, P.F., de Vries, W.R. & Eikelboom, B.C. (2002) Detection and treatment of claudication due to functional iliac obstruction in top endurance athletes: a prospective study. *Lancet*, 359, 466–473.
- Schep, G., Kaandorp, D.W., Bender, M.H., Weerdenburg, H., van Engeland, S. & Wijn, P.F. (2001) Magnetic resonance angiography used to detect kinking in the iliac arteries in endurance athletes with claudication. *Physiological Measurement*, 22, 475–487.
- Schep, G., Schmikli, S.L., Bender, M.H., Mosterd, W.L., Hammacher, E.R. & Wijn, P.F. (2002) Recognising vascular causes of leg complaints in endurance athletes. Part 1: validation of a decision algorithm. *International Journal of Sports Medicine*, 23, 313–321.
- Shechtman, O. (2013) The coefficient of variation as an index of measurement reliability. In: *Methods of clinical epidemiology*. Berlin, DE: Springer, pp. 39–49.
- Spencer, M.D., Murias, J.M., Lamb, H.P., Kowalchuk, J.M. & Paterson, D.H. (2011) Are the parameters of VO₂, heart rate and muscle deoxygenation kinetics affected by serial moderate-intensity exercise transitions in a single day? *European Journal of Applied Physiology*, 111, 591–600.
- Synek, V. (2008) Evaluation of the standard deviation from duplicate results. *Accreditation and Quality Assurance*, 13, 335–337.
- Team, R.C. (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
- Thiel, C., Vogt, L., Himmelreich, H., Hübscher, M. & Banzer, W. (2011) Reproducibility of muscle oxygen saturation. *International Journal of Sports Medicine*, 32, 277–280.
- Ubbink, D.T. & Koopman, B. (2006) Near-infrared spectroscopy in the routine diagnostic work-up of patients with leg ischaemia. *European Journal of Vascular and Endovascular Surgery*, 31, 394–400.
- Van Beekvelt, M.C., Colier, W.N., Wevers, R.A. & Van Engelen, B.G. (2001) Performance of near-infrared spectroscopy in measuring local O₂ consumption and blood flow in skeletal muscle. *Journal of Applied Physiology (1985)*, 90, 511–519.
- van Hooff, M., Schep, G., Bender, M., Scheltinga, M. & Savelberg, H. (2020) Sport-related femoral artery occlusion detected by near-infrared spectroscopy and pedal power measurements: a case report. *The Physician and Sportsmedicine*, 49, 241–244.
- van Hooff, M., Schep, G., Meijer, E., Bender, M. & Savelberg, H. (2018) Near-infrared spectroscopy is promising to detect iliac artery flow limitations in athletes: a pilot study. *Journal of Sports Medicine (Hindawi Publ Corp)*, 2018, 8965858.
- Weir, J.P. (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength and Conditioning Research*, 19, 231–240.

How to cite this article: van Hooff, M., Meijer, E.J., Scheltinga, M.R.M., Savelberg, H.H.C.M. & Schep, G. (2022) Test-retest reliability of skeletal muscle oxygenation measurement using near-infrared spectroscopy during exercise in patients with sport-related iliac artery flow limitation. *Clinical Physiology and Functional Imaging*, 42, 114–126.

<https://doi.org/10.1111/cpf.12738>