Near-infrared spectroscopy-derived muscle oxygen saturation on a 0% to 100% scale: reliability and validity of the Moxy Monitor

Andri Feldmann
Roger Schmitz
Daniel Erlacher

SPIE.

Near-infrared spectroscopy-derived muscle oxygen saturation on a 0% to 100% scale: reliability and validity of the Moxy Monitor

Andri Feldmann,[a,*] **Roger Schmitz,**[b] **and Daniel Erlacher**[a]
[a]University of Bern, Institute of Sport Science, Bern, Switzerland
[b]Fortiori Design LLC, Hutchinson, United States

Abstract. Near-infrared spectroscopy (NIRS) to monitor muscle oxygen saturation (SmO_2) is rapidly expanding into applied sports settings. However, the technology is limited due to its inability to convey quantifiable values. A test battery to assess reliability and validity of a 0% to 100% scale modeled by a commercially available NIRS device was established. This test battery applies a commonly used technique, the arterial occlusion method (AOM) to assess repeatability, reproducibility, and face validity. A total of 22 participants completed the test battery to scrutinize the 0% to 100% scale provided by the device. All participants underwent repeated AOM tests in passive and active conditions. The SmO_2 minimum and SmO_2 maximum values were obtained from the AOM and were used in the subsequent analysis. Repeatability and reproducibility were tested for equivalency and Bland–Altman plots were generated. Face validity was assessed by testing SmO_2 values against an *a priori* defined threshold for mixed venous blood during AOM response. The device exhibits an appropriately functional 0% to 100% scale that is reliable in terms of repeatability and reproducibility. Under the conditions applied in the test battery design, the device is considered valid for application in sports. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JBO.24.11.115001]

Keywords: muscle oxygen saturation; oxygenation; oximetry; physiological calibration.

Paper 190256RRR received Jul. 19, 2019; accepted for publication Oct. 24, 2019; published online Nov. 18, 2019.

1 Introduction

In recent years, muscle oxygen saturation (SmO_2) measured via near-infrared spectroscopy (NIRS) has developed into an affordable and readily available technology. The application of this technology in athletic settings is expanding, as fundamental questions are being addressed.[1] One of the major concerns in the application of NIRS in athletic settings is that common limitations of NIRS are not well understood and need to be properly addressed, while still allowing the technology to be utilized for its clear advantages.[2,3]

The problem associated with using NIRS to measure quantifiable values has been the unknown path length problem in the modified Beer–Lambert method. Without knowing the photon path length, it is impossible to derive quantifiable values from the returning NIRS signal. This has the effect that NIRS output is in relative values, most often expressed as arbitrary units of hemoglobin (Hb) and myoglobin (Mb). NIRS cannot differentiate between oxyhemoglobin (O_2Hb) and oxymyoglobin (O_2Mb) or deoxyhemoglobin (HHb) and deoxymyoglobin (HMb)[4] and therefore for the remainder of this paper the terms O_2Hb and HHb will be the sum of O_2Hb and O_2Mb and HHb and HMb. Because these arbitrary units are relative in nature, direct comparisons are difficult and usually limited to trends in the derived signal. To increase the robustness of the relative values, it is often recommended to use a saturation in a percentage using the following equation:[5,6]

$$SmO_2 = \frac{O_2Hb}{O_2Hb + HHb} \times 100.$$

Saturation takes into consideration the relative change of total hemoglobin (tHb) and the interaction between O_2Hb and HHb. The result of this type of saturation equation is SmO_2, as mentioned earlier, or an often-applied tissue oxygenation index (TOI). It is important to denote the difference between the two, as they are often used interchangeably. In SmO_2, the m indicates that the saturation is intended be isolated to the muscle layer. In TOI, the *T* indicates the measurement is an average of all tissues under the sensor. Then, term saturation refers to the availability and functionality of a 0% to 100% scale, whereas an index generally refers to a measurement ratio to be compared to a fixed standard. If a parameter includes the term saturation, it should be reliable and valid in terms of 0% to 100% scale. However, large manufacturer differences, including inter-optode distance, number of wavelengths used, spectral width, and algorithmic calculations, all raise questions about the scaling of SmO_2 and TOI.[7–9] These differences focus around interindividual and intersite (muscle site selection) variation in results.[10,11] When choosing a new muscle site or changing test participant altogether, the optical properties measured by the NIRS device changes. If the NIRS device cannot adjust for these changes in specific measurement site properties, a large variation is expected and seen. Statistical normalization approaches, such as a physiological calibration, are then often used to address these problems.[7,9,12]

The situation becomes even more complicated as no consensus "gold standard" for NIRS-derived values exists to compare against. Certain studies do refer to gold standard comparisons, looking at alternate measurement techniques to validate NIRS.

---

*Address all correspondence to Andri Feldmann, E-mail: andri.feldmann@ispw.unibe.ch

Invasive experiments, such as isolated hindlimb experiments with animal subjects[13,14] or venous blood draw in humans performing exercise protocols,[15] show good results. Other publications comparing NIRS measurements to invasive measures of blood oxygenation levels demonstrate lower levels of success.[16,17] Noninvasive phosphorus magnetic resonance spectroscopy and NIRS comparisons[18] also show good results. Interesting, in particular, is the comparison between functional magnetic resonance imaging (fMRI), which is noted by Cui et al.[19] as a gold standard in human brain imaging and NIRS. FMRI as blood oxygen-level-dependent imaging requires the user to make almost near-identical assumptions about the measurement technique being used as NIRS does.[6] When debating a specific gold standard to which NIRS-derived SmO_2 should be compared, an obvious lack of authority exists as the term is not mentioned in reviews,[6,20,21] with appropriate hesitation. Alternate methods are needed to address the question of validity of SmO_2 measurements on a 0% to 100% scale.

A technique used for addressing the difficulty of relative values provided by NIRS, including the interindividual and intersite measurement dilemma, is the normalization process of physiological calibration.[20,21] This technique applies the arterial occlusion method (AOM) to identify the minimum and maximum signals of the NIRS device in question. The AOM functions by applying a suprasystolic cuff to a chosen limb for a 5- to 6-min period, which results in a linear change in the NIRS-derived signal for O_2Hb to a minimum signal point identified by a plateau. Upon release, the phenomenon of hyperemia[22] results in a return of the NIRS signal to a maximum point; the opposite is true for HHb. The minimum signal is a maximally deoxygenated state defined by a disappearance of O_2Hb in relation to the sum of O_2Hb and HHb, and maximum signal is a maximally oxygenated state and the disappearance of HHb to the sum O_2Hb and HHb. Having identified the minimum and maximum points, these can then be set at 0% to 100% as a physiological calibration. This process generates an individual and robust scale for further testing. Considering the debate around a true gold standard, this type of physiological calibration offers a functional test to assess and compare NIRS devices and SmO_2 scaling. In athletic settings and other possible applications of NIRS, using a standard AOM to calibrate the NIRS signal for each measurement site is not a feasible approach. Therefore, an NIRS device that provides a reliable and valid 0% to 100% scale with reasonable accuracy would greatly enhance NIRS usability.

To address the question of validation and reliability of an NIRS device, three concepts should be applied:

1. Repeatability[23] refers to closeness of agreement between repeated measurements made on the same participant under identical conditions.

2. Reproducibility[23] refers to the closeness of agreement between repeated measurements made under changing conditions. Under this rubric, three underlying concepts can be addressed: the intersite and interindividual differences discussed early, as well as change in muscle activation from a passive AOM test to an AOM test under active conditions or interactivation differences. These approaches have been examined using NIRS,[24,25] including a study by Lacroix et al.[26] using the AOM to investigate the NIRS reproducibility

during brachial artery occlusions yielding a high degree of reproducibility.

3. Face validity[27] refers to the reasonable expectation of measurements taken based on selected criteria. This last approach presumes to address the question of physiological validity, which is difficult to answer using NIRS. A possible criterion to setup thresholds to test against is a measure of venous oxygen saturation (SvO_2).

This paper applied the three identified concepts above in a specific test battery to evaluate the performance of SmO_2 on a scale of 0% to 100% provided by a commercially available NIRS device, the Moxy Monitor.

2 Methods

2.1 Participants

A total of 22 participants, 11 males and 11 females, took part in the study {age 21.8 ± 1.6 years; height 173.3 ± 9.9 cm; weight 67.0 ± 10.7 kg [mean $\pm$ standard deviation(SD)]}. All participants were Caucasian, in good health, nonsmokers, and unmedicated. Skinfold measures were taken using skinfold calibers to assess adipose tissue thickness (ATT) at the four measurement sites: vastus lateralis (VL) 12.5 ± 5.1 mm; rectus femoris (RF) 13.9 ± 5.3 mm; vastus medialis (VM) 13.3 ± 6.1 mm; gastrocnemius (G) 11.1 ± 5.9 mm. The participants were informed of the study design and the physical tasks ahead of time and written informed consent was obtained in advance. The study was carried out in accordance with the 1964 Declaration of Helsinki. The protocol was approved by the ethics committee of the local Faculty of Human Sciences.

2.2 Moxy Monitor

The Moxy Monitor (Fortiori Design LLC) is an NIRS device that propagates to provide an *a priori* 0% to 100% scale with accuracy useful for sports science applications. The device measures the amount of light reaching two detectors from one emitter at four wavelengths in a diffuse reflectance configuration for a total of eight measurements. The device detectors are spaced at 12.5 and 25 mm from the emitter. The default sampling rate cycles through the four wavelengths 80 times every 2 s and averages out the readings for an output rate of 0.5 Hz. As the device in focus, which propagates to provide an *a priori* 0% to 100% scale with reasonable accuracy, a clearer picture of the technical process involved to isolate and investigate the muscle layer and generate an SmO_2 output is provided. This should highlight the process with which this continuous wave device is able to overcome the path length problem and return absolute saturation values. However, perhaps of greater importance in determination of absolute saturation values and a focus of this paper is that the data output acquired through standardized experimentation should be assessed to determine reliability and validity, rather than technical approaches to the path length problem.

The measurement algorithm uses four steps to overcome the unknown path length problem:

1. The device applies a Monte Carlo model[28,29] to generate a large set of optical rays over the full measurement spectrum that travel from the emitter to the detectors through the predetermined tissue layers

consisting of epidermis, dermis, adipose, and muscle. The model uses published values[30–33] for scattering in these tissue types. The ray data include the path length in each layer and the model is run for numerous different ATT layers.

2. A data smoothing application is applied to reduce the effects of Monte Carlo statistical errors.

3. A matrix of expected detector measurements is generated from the ray trace data based on tissue optical properties that are expected to be encountered when measuring athletes, including the expected ranges of SmO_2 and tHb. This uses published values[30–33] for the absorbance of the chromophores that are modeled in each tissue layer.

4. A numerical solving and interpolating algorithm that compares the eight actual diffuse reflectance measurement matrices from step 3 to determine the optical properties (i.e., SmO_2 and tHb) of the muscle layer is applied.

The following equation shows how steps 1 to 3 are used to generate the matrix of expected detector measurements by applying the Beer–Lambert relationship to the Monte Carlo ray trace data:

$$I = f q \sum_{\lambda=\text{low}}^{\lambda=\text{high}} \sum_{i=1}^{m_\lambda} \frac{S_\lambda j_{0,\lambda,\text{samp}} \Delta\lambda}{n_\lambda}$$
$$\times e^{-(\mu_{A,e,\lambda} L_{e,\lambda,i} + \mu_{A,d,\lambda} L_{d,\lambda,i} + \mu_{A,a,\lambda} L_{a,\lambda,i} + \mu_{A,m,\lambda} L_{m,\lambda,i})},$$

where I is the total intensity of the optical detector output; f and q are the scaling parameters for wavelength-independent factors such as the optical coupling efficiency to the tissue and the LED brightness; the first summation is over the wavelength range of the light source; the second summation is over all rays that were traced to reach the detector; S_λ is the spectral sensitivity of the detector; $j_{0,\lambda} \Delta\lambda / n_\lambda$ is the initial power in each ray; μ_A is the absorption coefficient, which is the sum of the absorption coefficients of all relevant chromophores; L is the total path length in that layer of the i'th ray; and the subscripts e, d, a, and m refer to the tissue layers of epidermis, dermis, adipose, and muscle, respectively.

There are several important factors in this algorithm that attempt to overcome the limitations of the traditional modified Beer–Lambert techniques.

1. The Monte Carlo model accommodates the wavelength-dependent scattering differences across the measurement spectrum. The model returns the path length for each ray in each of the tissue layers to overcome the unknown path length problem of the Beer–Lambert law.

2. The Monte Carlo model includes the effects of an unknown ATT by modeling a range of ATT. A different set of ray trace data is used for each ATT in generating the matrix of expected detector measurements.

3. More subtly, the effective path length (EPL), even for a fixed set of traced rays with a distribution of path lengths, is dependent on the absorbance. In the limit of

absorbance approaching infinity, the EPL approaches the shortest path lengths in the distribution. In the limit of absorbance approaching zero, the EPL approaches the average path length of the distribution. The model overcomes this complexity by using the full ray trace set for all detector measurement predictions.

4. The model includes confounding factors, such as melanin, water, LED spectral width, and varying detector spectral sensitivity, which accommodate their presence in the measurement and allow the wavelength selection and solving algorithms to be designed to minimize sensitivity to these factors.

5. The algorithm includes the LED spectral sensitivity to temperature, which is accommodated by a temperature sensor in the device.

The Moxy Monitor has been compared with alternate NIRS device and evaluated for reliability and validity in previously published papers.[7,34]

2.3 Near-Infrared Spectroscopy Measurement

The sensors were mounted on four muscles, ensuring that minimal spacing between interdevice receiver and detector of 10 cm was maintained to avoid interference. The first sensor was placed on the VL at two-thirds between anterior superior iliac spine and the lateral side of the patella. The second sensor was placed on the RF half way between the anterior superior iliac spine and the top part of the patella. The third sensor was placed on the VM four-fifth down along the line of the anterior superior iliac spine and anterior border of the medial ligament. The final sensor was placed on the lateral head of the G at one-third of the way between the head of the fibula and the heel. All locations are as recommended by the SENIAM project[35] for electromyography measurements. The emitter and detectors were aligned in the direction of muscle fibers, and body hair was removed from the sensor sites. The sensors were fixed in place using medical adhesive tape (Hypafix; BSN Medical, DE) and were then covered with the compatible commercially available light shield to eliminate possible ambient light intrusion.

2.4 Procedure

A series of tests were selected to address the questions of reliability and validity on a 0% to 100% scale generated by the selected NIRS device. The test battery was designed to examine repeatability, reproducibility, and face validity. To limit the physiological variation and ensure a stable and repeatable environment, all tests used the AOM. For the experimental procedure, participants came into the lab for two session with 1-week separation between each session, as shown in Fig. 1.

2.4.1 Arterial occlusion method

The AOM was conducted using a pneumatic tourniquet (Rudolf Riester GmbH, DE) with thigh cuff dimensions of 96×13 cm inflated to >300 mm Hg. The tourniquet was suited on the right leg of all the participants. Prior to every test, all participants were asked to refrain from strenuous physical activity 24 h prior, to refrain from alcohol consumption and smoking 24 h prior to the experiment, and to maintain individual diet routine. The maximally deoxygenated state plateau identified as SmO_2

	Session 1	Session 2
Repeatability	AOM passive trial 1	AOM passive trial 2
Reproducability: a. Intersite b. Interindividual c. Interactivation	AOM passive trial 1 Leg extension 1RM setup trial	1 Week → AOM passive trial 2 AOM active trial
Face Validity	AOM passive trial 1	AOM passive trial 2

Fig. 1 Experimental procedure and aim for each participant. The test battery involved two AOM test sessions including two passive AOM trials and one active AOM trial.

minimum (SmO_2 min) was determined by the average of the final 20 s or 10 data points of the AOM, as long as this met the condition of a visual plateau. The maximally oxygenated state identified as SmO_2 maximum (SmO_2 max) was determined as the peak SmO_2 output average over 10 s or 5 data points following the end of the AOM as a result of the hyperemic effect.

2.4.2 Passive trials

Sensors were placed on the VL, VM, RF, and G and the participants assumed a lying supine position. Participants assumed the lying position for 5 min prior to data collection. The data collection started with 60 s of data collection for a baseline measurement, and after 60 s the pneumatic tourniquet was rapidly inflated. The pneumatic tourniquet remained inflated and pressure controlled for the 6 min to find the SmO_2 min plateau. The quality of the arterial occlusion was controlled through pulse oximeter and pulse palpation of the lower leg. After 6 min, the pneumatic tourniquet was released, and an additional 3 min of measurement took place to assess the hyperemic response and to find the SmO_2 max value.

2.4.3 Active trial

Each participant came in for an initial setup session to determine 1 repetition-maximum (1-RM). Participants executed a series of maximum effort leg extension trials on a leg extension machine (Schnell GmbH, DE). The best trial was taken as their estimated 1-RM. In the active AOM session, each participant was again suited with the pneumatic tourniquet and the sensors placed on the activity-recruited muscles VL, VM, and RF. The participants were then positioned in the knee extension machine and remained in the sitting position for 5 min prior to data collection. The data collection started with 60 s of data collection for a baseline measurement and, after the 60 s, the pneumatic tourniquet was rapidly inflated. The participants then executed continuous leg extension repetitions at 40 rpm at 5% of 1-RM until exhaustion. The pneumatic tourniquet remained inflated and pressure controlled as long as activity took place to identify SmO_2 min. Following exhaustion, the pneumatic tourniquet was released, and an additional 3 min of measurement took place to assess the hyperemic response and to find the SmO_2 max value.

2.5 Statistical Analysis

Owing to the confounding effects of ATT,[36] all measurement sites with ATT greater than 60% of the emitter-detector distance (15.0 mm) were removed from the analysis. The Shapiro–Wilk

test was selected for all data sets to test for normal distribution because of the small sample sizes used in the study. Statistical computations were performed using Microsoft Excel for Windows (Version 16.0.4738.1000) and MathWorks Matlab for Windows (Version 9.3.0.713570 R2017b). Equivalency testing was used as the groundwork for statistical procedure based on the confidence interval (CI) comparisons and *a priori* equivalency intervals (EIs) to determine statistical equivalency and statistical difference in accordance with studies by Lakens[37] and Cumming and Finch.[38] The *a priori* determined EI for SmO_2 was set at ±5%.

2.5.1 Repeatability, interindividual, and intersite reproducibility: passive trials

To assess device repeatability, a Bland–Altman plot[39] was constructed for the two extracted values of SmO_2 max and SmO_2 min for all four repeated measurement sites: VL, VM, RF, and G. Upper and lower limits of agreement were set at 1.96 SD and 95% CIs were calculated. An EI was set at ±5%. Pearson's correlation coefficients were calculated and tested for significance to assess the relationship between mean and mean difference. For interindividual and intersite reproducibility, all means and mean differences were plotted for all muscle sites with 90% CI and 95% CI for equivalency testing.

2.5.2 Interindividual, intersite, and interactivation reproducibility: active trial

The attainable values for SmO_2 min and SmO_2 max for the active and passive conditions were displayed in a Bland–Altman plot[39] to determine the interactivation reproducibility for VL, VM, and RF. Upper and lower limits of agreement were set at 1.96 SD and 95% CI were calculated. The same EI was set at ±5%. Pearson's correlation coefficients were calculated and tested for significance to assess the relationship between mean and mean difference. For interindividual and intersite reproducibility, all means and mean differences were plotted for all muscle sites with 90% CI and 95% CI for equivalency testing. For the calculations, the mean of the Moxy passive trials was used against the corresponding active trial and therefore proper adjustments to SD were made as proposed by Bland and Altman.[40]

2.5.3 Face validity: 0% to 100% scale

To determine face validity, a comparison was made between SmO_2 results from the AOM passive Moxy trials and documented values for SvO_2 post AOM tests. Hamaoka et al.'s

results[41] show an end AOM value for SvO_2 of $26.2 \pm 6.4\%$. Langham et al.'s results[42] show an end AOM hyperemic value for SvO_2 of $85 \pm 8.0\%$. These values were used to establish *a priori* thresholds for SmO_2 max and SmO_2 min for comparison with the results of the AOM passive trials. Means for all muscle sites SmO_2 max and SmO_2 min with 90% CI and 95% CI were plotted against the *a priori* thresholds to assess difference.

3 Results

All muscle sites display the expected linear decrease in SmO_2 with the application of the suprasystolic cuff to a minimum plateau and upon release a hyperemic response in both passive and active conditions (see Fig. 2). The 0% to 100% scale showed a good dynamic range over all muscle sites during passive conditions with M_{range} of $67.9\% \pm 9.9$ ($M_{min} = 10.1\% \pm 5.7$; $M_{max} = 78.1\% \pm 6.0$), as shown in Fig. 3.

3.1 Repeatability, Interindividual, and Intersite Reproducibility: Passive Trials

Looking at the mean difference of SmO_2 min, repeatability during passive trials in all muscle sites showed no statistical difference between passive trials 1 and 2 and all can be considered statistically equivalent using EI (see Fig. 4). For SmO_2 max,

repeatability during passive trials all muscle sites showed no statistical difference between passive trials 1 and 2, but only the VL and G sites can be considered statistically equivalent (Fig. 4). Bland–Altman plot analysis was used in addition to assess equality on a case-to-case basis. All Bland–Altman results were considered to show suitable agreement between passive trials 1 and 2 (Fig. 5). All muscles show no systemic bias between trials 1 and 2, as the line of equality is clearly within the EI. The data show that the SmO_2 max hyperemia has a greater degree of variation than SmO_2 min as a result of the AOM (Fig. 5). However, none of the plots shows significant relationship between mean and difference in both the max and min responses following a Pearson product–moment correlation. Looking at the mean of SmO_2 min and SmO_2 max for interindividual and intersite reproducibility during passive trials, no statistical difference between the obtained values can be discerned (Fig. 2).

3.2 Interindividual, Intersite, and Interactivation Reproducibility: Active Trial

Looking at reproducibility between active trials and passive trials, mean difference comparisons show no difference and statistical equivalency between VL SmO_2 min and VM SmO_2 min
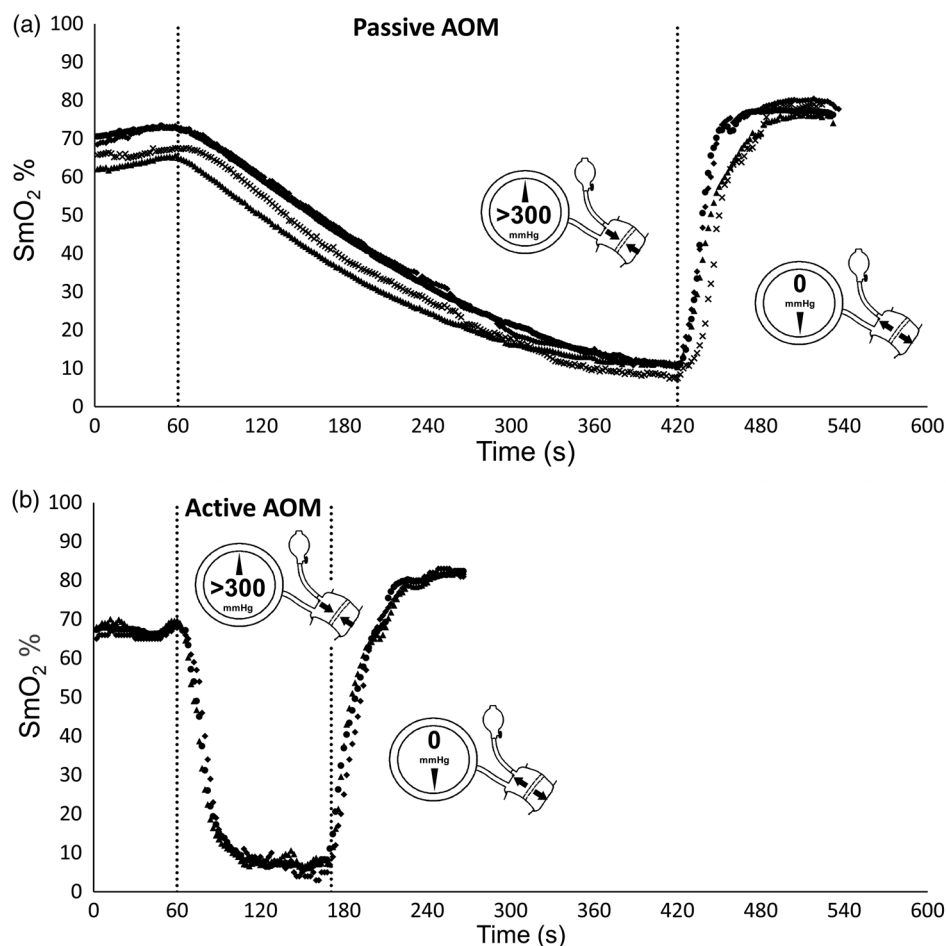

Fig. 2 Predicated linear decrease in SmO_2 as a result of the AOM and time during (a) passive and (b) active conditions, with minimum value plateau and hyperemic response on cuff release. Dotted lines indicate start and stop of the AOM. Mean VL (○); mean VM (◊); mean RF (△); mean G (x).
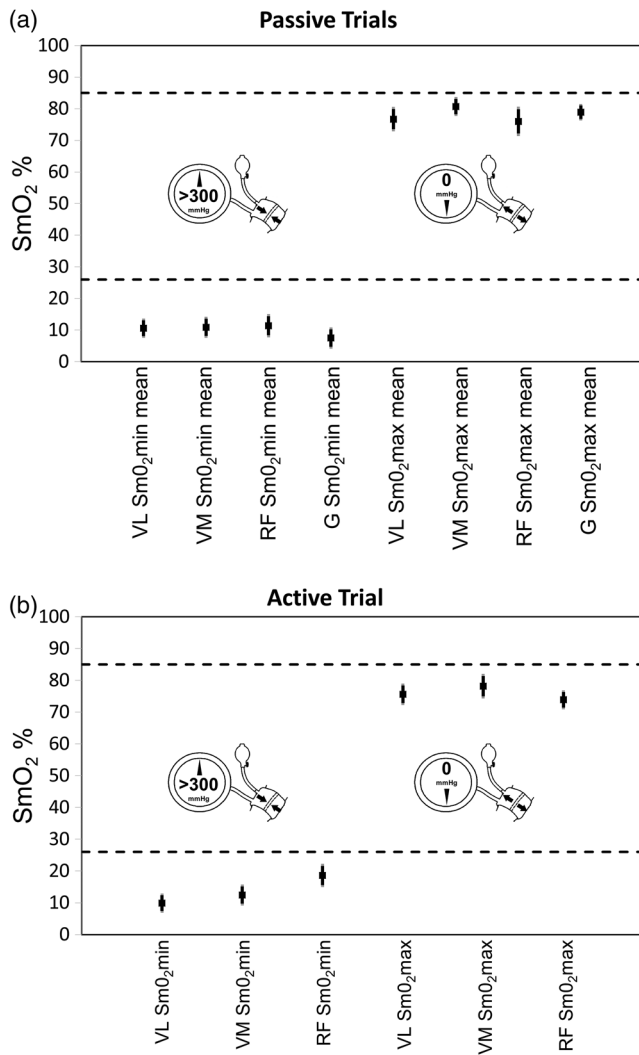
Fig. 3 Mean (black squares) and 90% CI (thick darker vertical lines) and 95% CI (thin lighter vertical lines) for SmO$_2$ min and SmO$_2$ max for (b) active and (a) passive conditions for all muscle sites: VL, VM, RF, and G. Dashed lines indicate the *a priori* determined thresholds of SvO$_2$ for assessment of face validity.

using the EI (see Fig. 4). The same is true for VL SmO$_2$ max. RF SmO$_2$ max and VM SmO$_2$ max showed no difference but are not statistical equal (see Fig. 4). RF SmO$_2$ min was not statistically equivalent and different at the 95% CI. The results of SmO$_2$ max and SmO$_2$ min on a Bland–Altman plot for the three examined quadriceps muscle sites show acceptable equivalency for VL and VM but not for RF between passive and active trials (see Fig. 6). As with the comparison between passive trials, the active–passive trials show a greater degree of variation in the SmO$_2$ max hyperemia results. Unlike the passive trial comparison, the active-passive trials show a significant relationship between mean and difference for SmO$_2$ min in a Pearson product-moment correlation with a tendency toward higher values for SmO$_2$ min for the active condition for VL SmO$_2$ min ($r = -0.5251$, $n = 15$, $p = 0.044$) and RF SmO$_2$ min [$r = -0.7071$, $n = 13$, $p = 0.006$ (see Fig. 6)]. Looking at the mean of SmO$_2$ min and SmO$_2$ max for interindividual and intersite reproducibility, during active and passive trials only RF SmO$_2$ min during the active trial shows potential difference in means (see Fig. 2).

3.3 Face Validity: 0% to 100% Scale

The *a priori* determined thresholds for SmO$_2$ min and SmO$_2$ max were tested against the mean of trials 1 and 2. For each muscle site, the passive trial mean was plotted with 90% CI and 95% CI and assessed against the *a priori* thresholds. With the exception of RF SmO$_2$ min, during active conditions all means for SmO$_2$ min and SmO$_2$ max lie below the *a priori* thresholds as predicted (see Fig. 2). RF SmO$_2$ min crosses the threshold at the 95% CI.

3.4 Effect of Adipose Tissue Thickness

Because of the well-documented effects of ATT on NIRS signal and consequently on SmO$_2$, the effect of ATT on SmO$_2$ min was plotted for each muscle site using all 22 participant's data (see Fig. 7). Linear and segmented regressions were calculated for each muscle site. Clearly, a relationship exists between ATT and SmO$_2$ min, as shown by the linear regression for all muscle sites: VL ($r = 0.844$, $n = 22$, $p < 0.001$); VM ($r = 0.901$, $n = 22$, $p < 0.001$); RF ($r = 0.872$, $n = 22$, $p < 0.001$); and G ($r = 0.912$, $n = 22$, $p < 0.001$). When using a segmented regression, all muscle sites identify a single knot at 15.37 ± 1.52 mm to optimize the linear regression (see Fig. 7). The results indicate to a certain degree an effectiveness in maximizing the sensitivity of the NIRS-derived SmO$_2$ signal as long as the ATT thickness remains within the recommended penetration depth threshold of 15 mm. SmO$_2$ values obtained over the ATT threshold should be considered suspect.

4 Discussion

The purpose of this experiment was, first, to propose a test battery to investigate reliability and validity of SmO$_2$ on 0% to 100% scale and, second, to apply this test battery to readily available NIRS device. The argument presented makes some assumptions, the first of which is that the process of a physiological calibration through AOM provides reliable and functional information about the range of NIRS-derived oxygenation signals at an individual level. This being true, a reliable and valid 0% to 100% scale can be scrutinized using the AOM. The 0% to 100% scale tested would need to show acceptable results for repeatability, reproducibility, and face validity, and therefore this test battery sets the groundwork to determine device functionality. In this experiment, all participants showed good repeatability and reproducibility during the AOM tests using the Moxy Monitor.

To discuss validity, an inference was made, and therefore as a product of deductive reasoning the term face validity was considered appropriate. The first position was that muscle tissue during activity or occlusion situations would represent the highest metabolic activity in comparison to other peripheral tissue being measured.[43] Therefore, when comparing NIRS-derived values for SmO$_2$ against invasive measures of SvO$_2$, the SmO$_2$ values cannot be higher than the measured value for SvO$_2$. SmO$_2$ should be lower, as SvO$_2$ is a combination of venous blood returning from all tissue layers, including adipose and skin tissue; this is the premise of venous blood contamination, which will be discussed later. While this does not establish validity of an NIRS-derived SmO$_2$ value, it does establish thresholds against which measured values can be tested and lends a useful SmO$_2$ range; the same argument is made by McManus et al.[7] The *a priori* SvO$_2$ thresholds applied to this study are
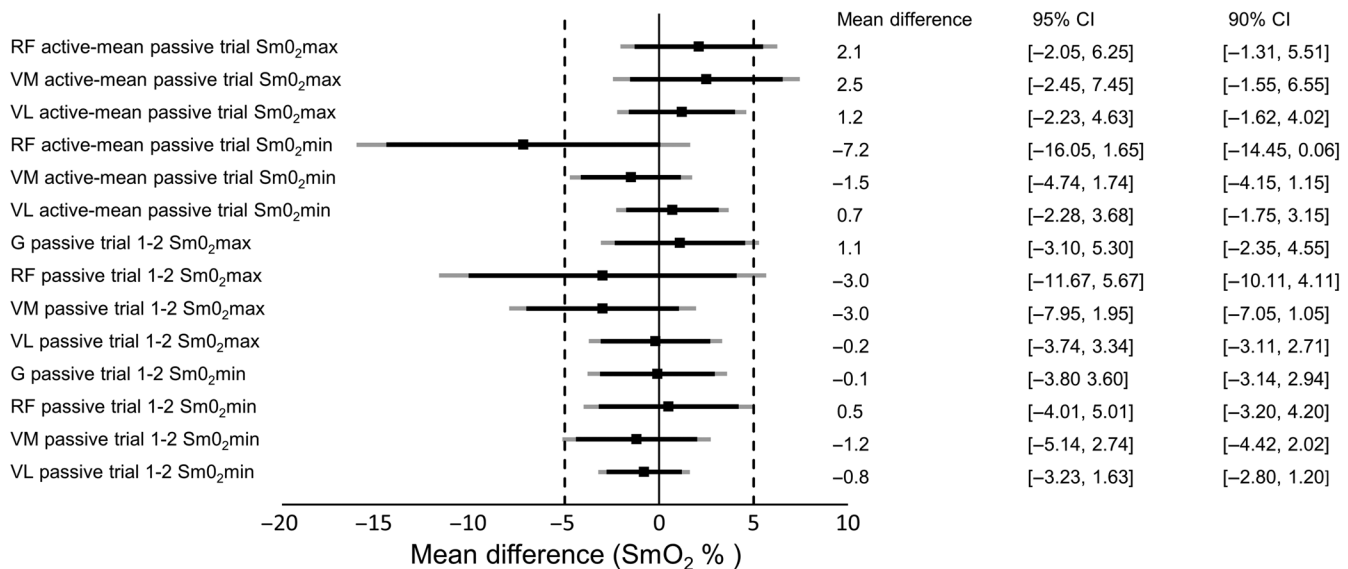
Fig. 4 Mean differences (black squares) and 90% CI (thick darker horizontal lines) and 95% CI (thin lighter horizontal lines). The mean differences for active and passive trials are the difference between passive trials mean and the active trial in question. *A priori* EI is set at $\pm 5\%$ (dashed lines). Solid line indicates line of equality.

drawn from SvO_2 data collected during AOM experiments and show the extent of oxygenation ranges in metabolic tissue.[41,42] This range has also been established during high-intensity exercise, during small muscle activation, and during full body exercise. Costes et al.[16] showed SvO_2 values of 17.8 ± 4.2 and 7.8 ± 2.3 in normoxia (N) and hypoxia (H), respectively, following 20 min of steady-state cycling at 80% of VO_2 max. Mancini et al.[15] and Macdonald et al.[17] showed SvO_2 ranges following small muscle and lower body isolated kicking exercises of $30.4 \pm 6.8\%$, and then 40.1 ± 2.2 (N) and 33.9 ± 1.8 (H). All three of these investigations included NIRS measurements. While for Mancini et al.[15] the NIRS data collected correlated with the SvO_2 data, for both Costes et al.[16] and Macdonald et al.[17] this was not the case under the normoxic condition. This confounds the relationship between NIRS-derived oxygenation signals and SvO_2. A reconciliation with these apparent contradictions involves the venous blood contamination problem discussed earlier.[6] Measured SvO_2 is a combination of blood from active skeletal muscles and skin and adipose tissue circulation. The venous blood contamination problem then results in increased or contradictory NIRS values, in comparison to SvO_2, because of increased contribution of lower metabolically active tissue to the NIRS signal as a result of, for example, skin blood flow increase for heat dissipation.[44,45] Advancements in the spatially resolved method has attempted to address this problem.[45] While this advancement has helped in the scaling of NIRS, when looking at device comparison data from this experiment, or alternative studies,[7–9] clearly a functional scale has not been determined. A device measuring SmO_2 must distinguish between measured tissue layers to isolate the muscle layer. This is particularly important because certain NIRS devices provide a TOI measurement rather than a SmO_2 measurement. This difference should be considered in the analysis and discussion of NIRS data, as the two terms should not be used interchangeably.

To further address and discuss this useful accuracy of SmO_2, assumptions about signal contribution need to be made. Under normal conditions, muscles receive near completely oxygenated arterial blood[8] and therefore it can be assumed changes in oxygenation reported by the NIRS signal can be attributed to changes in Hb and Mb oxygenation at the examined tissue level. The physical demands of the Beer–Lambert law exempt large blood vessels from contributing in a significant way to the NIRS-derived signal, as the concentration of absorbing chromophores is too large to return a signal.[6] This means that the NIRS-derived signal is mostly from smaller vessel contributions along the lines of arterioles, capillaries, and venules. How much of the signal is derived from which source is a discussion,[46–48] and a commonly used formula is equal parts contribution by all three-vessel systems[7] or the "one-third, one-third, one-third" model. This assumption led to the discussion by McManus et al.[7] that the range of SmO_2 by the Moxy Monitor is rather large. Interestingly, McManus et al.[7] apply the same logic using SvO_2 thresholds to discuss a potential physiological validity, just applying a different assumption of contribution. The paper does go on to discuss potential difference in NIRS signal as a result of muscle layer specialization—stressing the importance of terminology between for TOI and SmO_2. Nonetheless, the one-third, one-third, one-third model stands to be disputed. Boushel et al.[47] argue for a 10:20:70 ratio of arterial, capillary, and venous blood in the NIRS signal. An experiment by Poole and Mathieu-Costello[48] shows that >90% of total blood volume in the muscles is in the capillaries, which would then again return to the question of tissue layer isolation when talking about signal contribution. Clearly, the one-third, one-third, one-third model for signal contribution is questionable and an adequate model is up for debate. As this paper has no experimental claim to credit or discredit magnitude of signal contribution as was discussed, the assumption that a device advocating a 0% to 100% scale of reasonable accuracy in the form of physiological validity for SmO_2 should have results that are smaller than the measured values of SvO_2 remains intact as a matter of deduction.
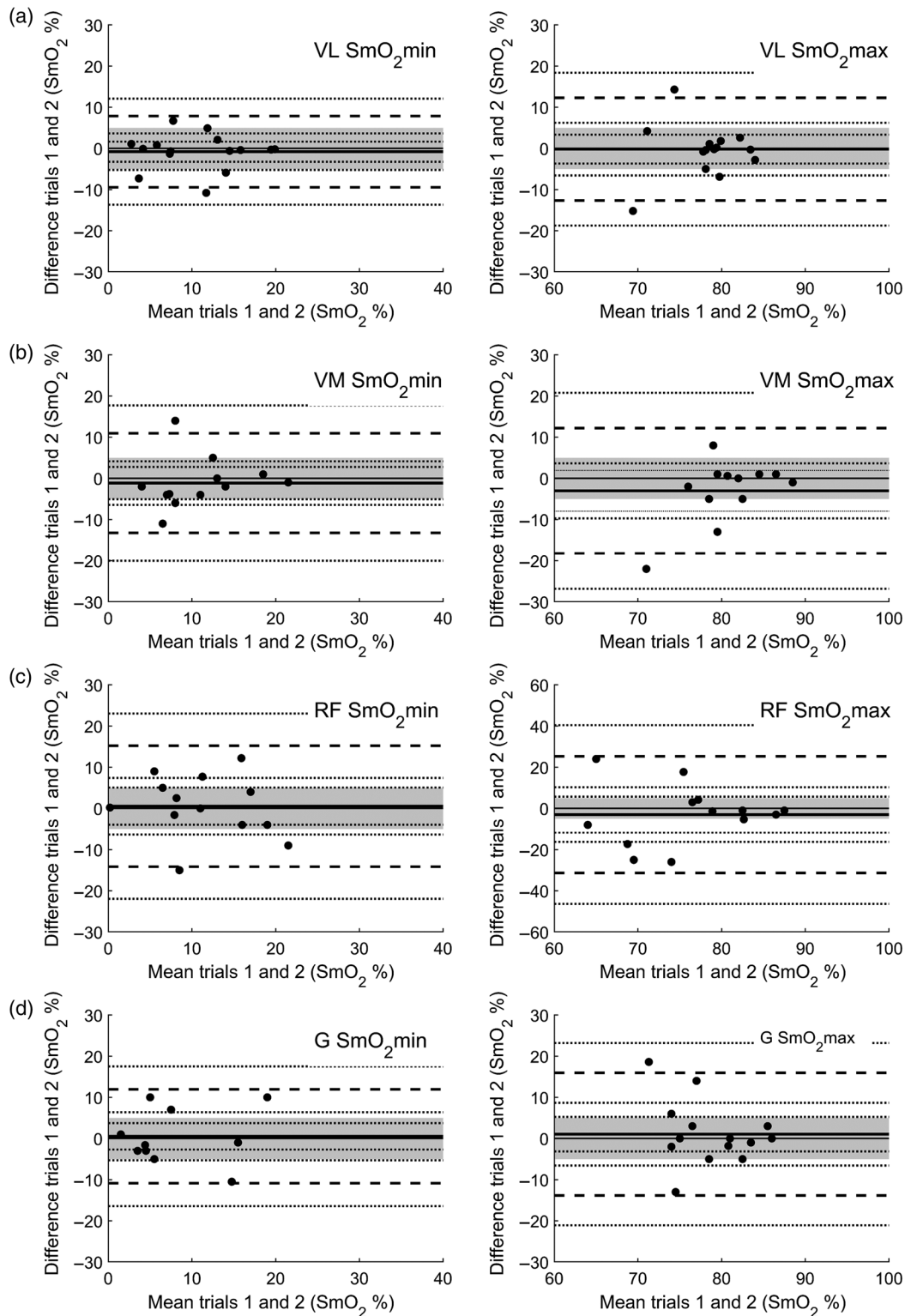
Fig. 5 Bland–Altman plots of passive trials 1 and 2 for SmO_2 min and SmO_2 max for (a) VL, (b) VM, (c) RF, and (d) G. *A priori* EI is set at $\pm 5\%$ (shaded area). The solid line identifies the mean bias (MB) and dashed lines identify the upper and lower limits of agreement at ± 1.96 SD. For both the MB and the limits of agreement, respectively, dotted lines represent 95% CI.

4.1 Limitations

The conducted experiment has its limitations, first and foremost in the assumptions that are repeatedly discussed. These assumptions stand to be refuted. The assumption of the SvO_2 thresholds relies on third-party data collections and therefore the question of comparison needs to be addressed. Both papers cited[41,42] use similar population pools and apply the same AOM to determine maximum and minimum values. For this reason, it was determined to be suitable to use these data to bridge the question
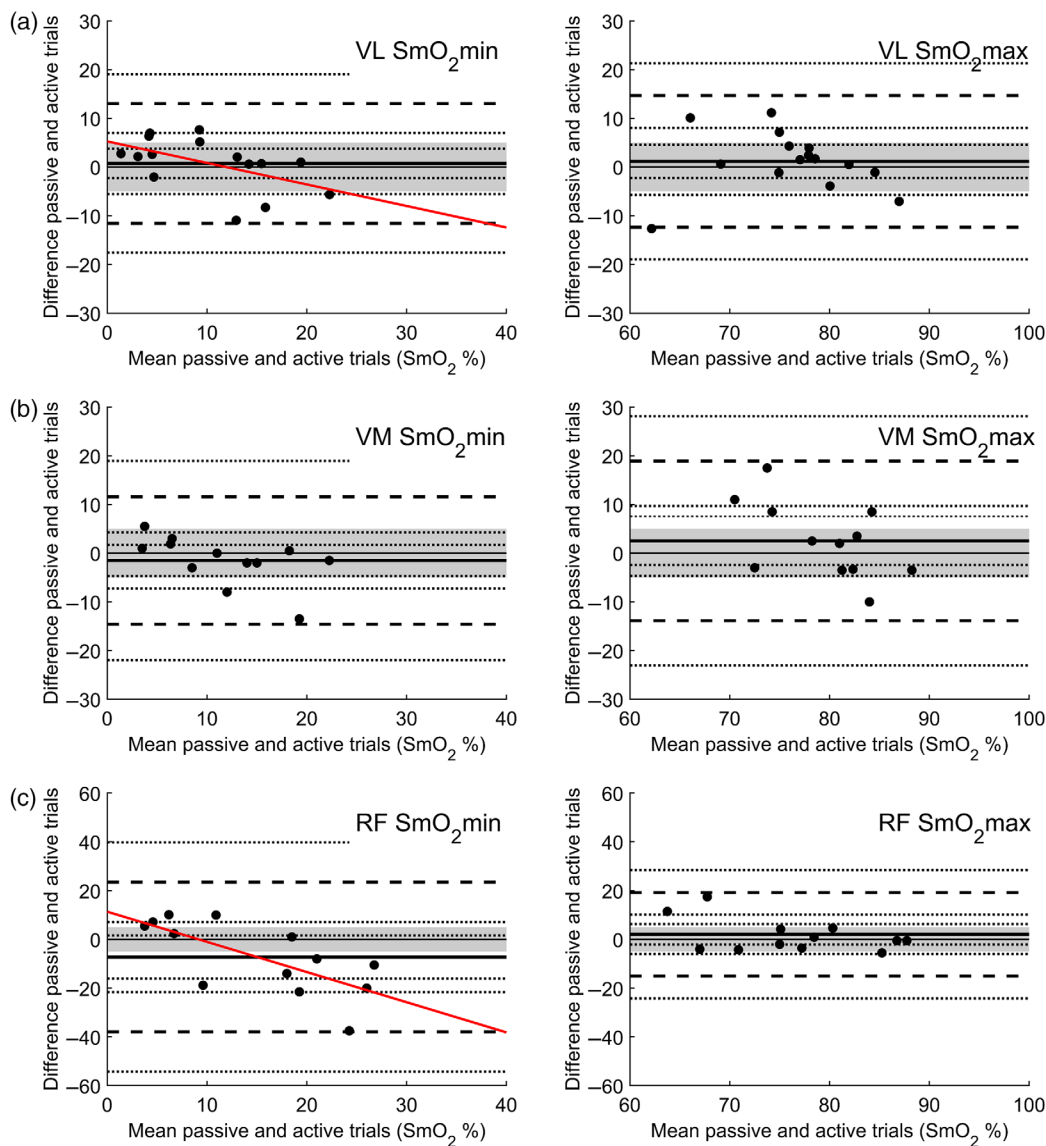
Fig. 6 Bland–Altman plots of passive trials mean and active trials for SmO$_2$ min and SmO$_2$ max for (a)VL, (b) VM, and (c) RF. *A priori* EI is set at ±5% (shaded area). The solid line identifies the MB and dashed lines identify the upper and lower limits of agreement at ±1.96 SD. For both the MB and the limits of agreement, respectively, dotted lines represent 95% CI. For RF SmO$_2$ min and VL SmO$_2$ min, a significant relationship was found between mean and mean difference (diagonal correlation line) at $R^2 = 0.2727$ and $R^2 = 0.4999$, respectively.

of physiological validity. It is highly recommended that this experiment is duplicated with venous blood sampling or alternative forms of physiological validation. The participant pool suffers from a large degree of homogeneity in terms of age, activity level, and melanin content (all participants were Caucasian). Further data need to be collected on differing demographic groups. ATT is a major concern for NIRS, as shown in this paper by the exclusion of measurement sites used in the analysis due to ATT. Caution is recommended when using the Moxy Monitor with ATT >15 mm. As identified in Sec. 4, arterial blood is always nearly completely oxygenated prior to the AOM. This study does not look at what would happen if you were to manipulate arterial oxygen saturation prior to the AOM. Generally, the SmO$_2$ min and SmO$_2$ max data appear to be consistent across muscle sites, individuals, and during muscularly active and passive conditions, there is some deviation in

the RF between active and passive conditions. This identifies further need for trials and investigation. This may be the result of the effect of position and occlusion cuff during the knee extensions, muscle requirement, and a greater degree of individual physiological variation during hyperemia. As can be seen in the Bland–Altman plots (see Fig. 6), the RF during the active conditions has a few substantial outliers. Outliers were not removed for transparency purposes. Finally, for the statistical analysis of SmO$_2$ min and SmO$_2$ max, averages were collected for the identifiable plateaus and limits. Considering the low sampling rate of the device, these calculations involved a small number of data points, which should be concerning, as they are subject to noise, as is any measurement. However, as pointed out in the device specifications, while it is correct that the sampling rate in terms of received data is 0.5 Hz, this output value is already a product of smoothing over 80 LED cycles. Therefore,
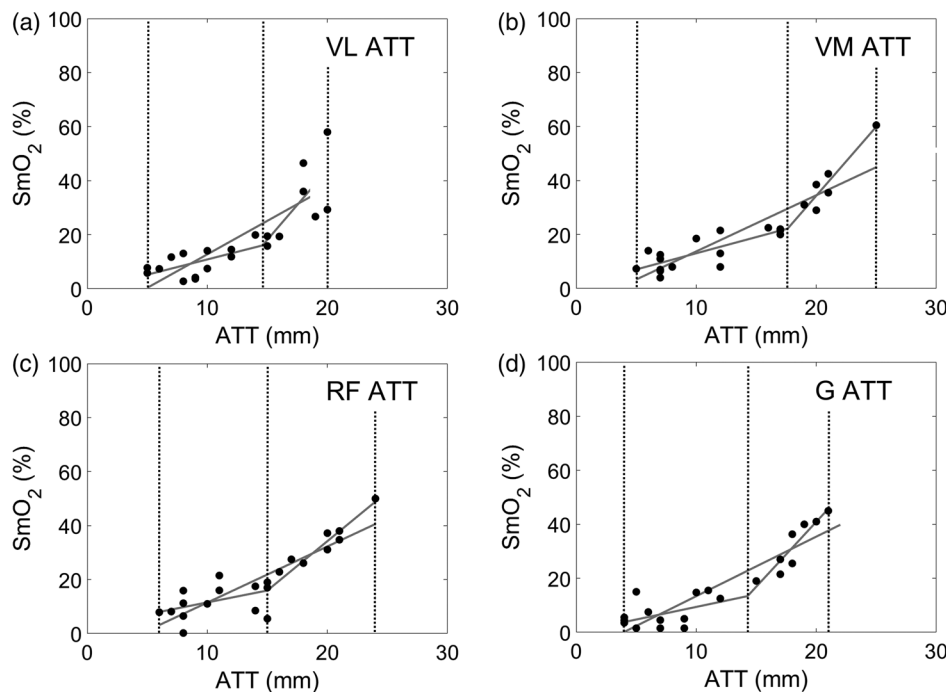
Fig. 7 Dashed lines indicate the linear model and segmented regression model with one knot point. Dotted lines indicate start and stop of analysis data points and the knot point. The optimal knot minimizes sum of squares residuals calculated at (a) VL 14.595 mm, $R^2 = 0.798$; (b) VM 17.613 mm, $R^2 = 0.928$; (c) RF 15.000 mm, $R^2 = 0.862$; and (d) G 14.270 mm, $R^2 = 0.907$.

the output information has already been, to a certain extent, controlled for noise as it is a product of a much larger sampling rate. Still, while considerations were made, time course components of the NIRS parameters were left out of the analysis because of the 0.5-Hz output rate.

5 Conclusion

The study illustrates that the retail NIRS device Moxy Monitor is valid and reliable under the conditions of the test battery used. The validity attributed to the device in this paper is a consequence of a series of assumptions, which should be viewed critically. While the repeatability and reproducibility comparison were successful and the functionality of NIRS to measure changes in SmO_2-related supply-and-demand parameters does not stand in question, the validity of a 0% to 100% scale is open for discussion. Nonetheless, this type of scaling is of great importance for research and athletic applications in order to compare and contrast data. For this reason, the authors recommend the use of functional scales of 0% to 100% that reflect to an acceptable degree physiological validity, in this case to the reference to SvO_2. In the absence of a gold standard, functional scales should be tested in the form of physiological calibration via AOM involving this type of testing battery, including tests of repeatability, reproducibly, and validity.

References

1. S. Perrey and M. Ferrari, "Muscle oximetry in sports science: a systematic review," *Sports Med.* **48**(3), 597–616 (2018).
2. D. P. Born et al., "Near-infrared spectroscopy: more accurate than heart rate for monitoring intensity in running in hilly terrain," *Int. J. Sports Physiol. Perform.* **12**(4) 440–447 (2017).
3. B. Jones, M. Dat, and C. E. Cooper, "Underwater near-infrared spectroscopy measurements of muscle oxygenation: laboratory validation and preliminary observations in swimmers and triathletes," *J. Biomed. Opt.* **19**(12), 127002 (2014).
4. F. F. Jöbsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science* **198**(4323), 1264–1267 (1977).
5. V. Quaresima and M. Ferrari, "Muscle oxygenation by near-infrared-based tissue oximeters," *J. Appl. Physiol.* **107**(1), 371–371 (2009).
6. B. Grassi and V. Quaresima, "Near-infrared spectroscopy and skeletal muscle oxidative function in vivo in health and disease: a review from an exercise physiology perspective," *J. Biomed. Opt.* **21**(9), 091313 (2016).
7. C. J. McManus, J. Collison, and C. E. Cooper, "Performance comparison of the MOXY and PortaMon near-infrared spectroscopy muscle oximeters at rest and during exercise," *J. Biomed. Opt.* **23**(1), 015007 (2018).
8. S. Hyttel-Sorensen et al., "Tissue oximetry: a comparison of mean values of regional tissue saturation, reproducibility and dynamic range of

four NIRS-instruments on the human forearm," *Biomed. Opt. Express* **2**(11), 3047–3057 (2011).

9. J.-H. Lee et al., "Comparison of two devices using near-infrared spectroscopy for the measurement of tissue oxygenation during a vascular occlusion test in healthy volunteers (INVOS® vs. InSpectra™)," *J. Clin. Monit. Comput.* **29**(2), 271–278 (2015).

10. H. Gómez et al., "Characterization of tissue oxygen saturation and the vascular occlusion test: influence of measurement sites, probe sizes and deflation thresholds," *Crit. Care* **13**(5), S3 (2009).

11. R. Bezemer et al., "Assessment of tissue oxygen saturation during a vascular occlusion test using near-infrared spectroscopy: the role of probe spacing and measurement site studied in healthy volunteers," *Crit. Care* **13**(5), S4 (2009).

12. T. Komiyama et al., "Comparison of two spatially resolved near-infrared photometers in the detection of tissue oxygen saturation: poor reliability at very low oxygen saturation," *Clin. Sci.* **101**, 715–718 (2001).

13. J. R. Wilson et al., "Noninvasive detection of skeletal muscle underperfusion with near-infrared spectroscopy in patients with heart failure," *Circulation* **80**, 1668–1674 (1989).

14. Y. Sun et al., "Muscle near-infrared spectroscopy signals versus venous blood hemoglobin oxygen saturation in skeletal muscle," *Med. Sci. Sports Exercise* **48**(10), 2013–2020 (2016).

15. D. M. Mancini et al., "Validation of near-infrared spectroscopy in humans," *J. Appl. Physiol.* **77**(6), 2740–2747 (1994).

16. F. Costes et al., "Comparison of muscle near-infrared spectroscopy and femoral blood gases during steady-state exercise in humans," *J. Appl. Physiol.* **80**(4), 1345–1350 (1996).

17. M. J. MacDonald et al., "Comparison of femoral blood gases and muscle near-infrared spectroscopy at exercise onset in humans," *J. Appl. Physiol.* **86**(2), 687–693 (1999).

18. T. E. Ryan et al., "A cross-validation of near-infrared spectroscopy measurements of skeletal muscle oxidative capacity with phosphorus magnetic resonance spectroscopy," *J. Appl. Physiol.* **115**(12), 1757–1766 (2013).

19. X. Cui et al., "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *Neuroimage* **54**(4), 2808–2821 (2011).

20. T. Hamaoka et al., "Near-infrared spectroscopy/imaging for monitoring muscle oxygenation and oxidative metabolism in healthy and diseased humans," *J. Biomed. Opt.* **12**(6), 062105 (2007).

21. T. Hamaoka et al., "The use of muscle near-infrared spectroscopy in sport, health and medical sciences: recent developments," *Philos. Trans. R. Soc. A* **369**(1955), 4591–4604 (2011).

22. C. L. Murrant, I. R. Lamb, and N. M. Novielli, "Capillary endothelial cells as coordinators of skeletal muscle blood flow during active hyperemia," *Microcirculation* **24**(3), e12348 (2017).

23. B. N. Taylor and C. E. Kuyatt, "Guidelines for evaluating and expressing the uncertainty of NIST measurement results," NIST 1297 (1994).

24. A. Adami et al., "Reproducibility of NIRS assessment of muscle oxidative capacity in smokers with and without COPD," *Respir. Physiol. Neurobiol.* **235**, 18–26 (2017).

25. M. Muthalib et al., "Reliability of near-infrared spectroscopy for measuring biceps brachii oxygenation during sustained and repeated isometric contractions," *J. Biomed. Opt.* **15**(1), 017008 (2010).

26. S. Lacroix et al., "Reproducibility of near-infrared spectroscopy parameters measured during brachial artery occlusion and reactive hyperemia in healthy men," *J. Biomed. Opt.* **17**(7), 077010 (2012).

27. F. J. Gravetter and L. B. Forzano, *Research Methods for the Behavioral Sciences*, 4th ed., Cengage Learning, Boston (2011).

28. A. A. A. Halim et al., "A review: functional near infrared spectroscopy evaluation in muscle tissues using Monte Carlo simulation," *AIP Conf. Proc.* **1963**(1), 020031 (2018).

29. Y. Yang et al., "Influence of a fat layer on the near infrared spectra of human muscle: quantitative analysis based on two-layered Monte Carlo simulations and phantom experiments," *Opt. Express* **13**(5), 1570–1579 (2005).

30. V. Tuchin, *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*, 2nd ed., SPIE Press, Bellingham, Washington, PM166 (2007).

31. C. R. Simpson et al., "Near-infrared optical properties of ex vivo human skin and subcutaneous tissues measured using the Monte Carlo inversion technique," *Phys. Med. Biol.* **43**(9), 2465–2478 (1998).

32. S. A. Prahl, "A compendium of tissue optical properties," http://omlc.ogi.edu/spectra/hemoglobin (2012).

33. S. A. Prahl, "Optical absorption of water," http://omlc.ogi.edu/spectra/water/ (2012).

34. E. M. Crum et al., "Validity and reliability of the Moxy oxygen monitor during incremental cycling exercise," *Eur. J. Sport Sci.* **17**(8), 1037–1043 (2017).

35. SENIAM, "SENIAM Project," http://www.seniam.org/ (2004).

36. M. C. Van Beekvelt et al., "Adipose tissue thickness affects in vivo quantitative near-IR spectroscopy in human skeletal muscle," *Clin. Sci.* **101**(1), 21–28 (2001).

37. D. Lakens, "Equivalence tests: a practical primer for t tests, correlations, and meta-analyses," *Soc. Psychol. Personal Sci.* **8**(4), 355–362 (2017).

38. G. Cumming and S. Finch, "Inference by eye; confidence intervals and how to read pictures of data," *Am. Psychol.* **60**(2), 170–180 (2005).

39. J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **327**(8476), 307–310 (1986).

40. J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods Med. Res.* **8**, 135–160 (1999).

41. T. Hamaoka et al., "Quantification of ischemic muscle deoxygenation by near infrared time-resolved spectroscopy," *J. Biomed. Opt.* **5**(1), 102–105 (2000).

42. M. C. Langham et al., "Evaluation of cuff-induced ischemia in the lower extremity by magnetic resonance oximetry," *J. Am. Coll. Cardiol.* **55**(6), 598–606 (2010).

43. Z. Wang et al., "Specific metabolic rates of major organs and tissues across adulthood: evaluation by mechanistic model of resting energy expenditure," *Am. J. Clin. Nutr.* **92**(6), 1369–1377 (2010).

44. B. Grassi et al., "Muscle oxygenation and pulmonary gas exchange kinetics during cycling exercise on-transitions in humans," *J. Appl. Physiol.* **95**(1), 149–158 (2003).

45. S. Koga et al., "Validation of a high-power, time-resolved, near-infrared spectroscopy system for measurement of superficial and deep muscle deoxygenation during exercise," *J. Appl. Physiol.* **118**(11), 1435–1442 (2015).

46. T. K. Tran et al., "Comparative analysis of NMR and NIRS measurements of intracellular PO_2 in human skeletal muscle," *Am. J. Physiol.* **276**(6), R1682–R1690 (1999).

47. R. Boushel et al., "Monitoring tissue oxygen availability with near infrared spectroscopy (NIRS) in health and disease," *Scand. J. Med. Sci. Sports* **11**(4), 213–222 (2001).

48. D. C. Poole and O. Mathieu-Costello, "Skeletal muscle capillary geometry: adaptation to chronic hypoxia," *Respir. Physiol.* **77**(1), 21–29 (1989).

Andri Feldmann received his BSc degree in human kinetics from the University of British Columbia and his MSc degree in sport science from the University of Bern. He was a codeveloper of the retail NIRS device the Moxy Monitor and a cofounder of the training consultancy and retail company Swinco AG. In 2017, the company was incorporated into the biotechnology company Idiag AG. In 2017, he started a PhD program at the University of Bern.

Roger Schmitz received his BS degree in mechanical engineering from Iowa State University and later studied biomedical optics at the University of Rochester. He developed medical NIRS devices for Hutchinson Technology Inc. He cofounded Fortiori Design LLC in 2010, which launched the Moxy muscle oxygen monitor NIRS device in 2013.

Daniel Erlacher is an associated professor of sport science at the University of Bern. His work focuses on the interaction between sleep and sport performance, e.g. the memory function of sleep. In recent years, he became interested in NIRS technology to gain insight into active muscles during sport and recovery.