Test–retest reliability of brain mitochondrial cytochrome-c-oxidase assessed by functional near-infrared spectroscopy

Lisa Holper
J. John Mann

Test–retest reliability of brain mitochondrial cytochrome-c-oxidase assessed by functional near-infrared spectroscopy

Lisa Holper* and J. John Mann
Columbia University, New York State Psychiatric Institute, Division of Molecular Imaging and Neuropathology, New York, New York, United States

Abstract. Functional near-infrared spectroscopy (fNIRS) is a noninvasive method for measuring in vivo both hemodynamic and mitochondrial metabolic activities in brain cortical structures. Although the test–retest reliability of the hemodynamic measures, such as reflected by oxygenated (HbO₂), deoxygenated (HHb) hemoglobin, and the tissue oxygenation index (TOI), has been previously reported to be good to excellent, the reliability of the metabolic signal indexed by oxidized cytochrome-c-oxidase (oxCCO) has not been reported. The present test–retest study compared the reliability of the metabolic and hemodynamic signals in 10 healthy participants undergoing hypox–hypercapnia challenges. The primary reliability measure was the intraclass correlation coefficient (ICC). Results of both hypo- and hypercapnia showed that the oxCCO signal (ICC = 0.876/0.757) had robust reliability comparable with that of the HbO₂ (ICC = 0.841/0.801), HHb (ICC = 0.804/0.571), and TOI (ICC = 0.574/0.614) signals. These findings show that the oxCCO signal can be assessed by fNIRS with comparable reliability to the hemodynamic measures. We discuss the results in light of current interest in a mitochondrial metabolic marker derived from fNIRS.

Keywords: cytochrome-c-oxidase; reliability; near-infrared spectroscopy.

Paper 170797RR received Dec. 8, 2017; accepted for publication Apr. 24, 2018; published online May 15, 2018.

1 Introduction

Functional near-infrared spectroscopy (fNIRS) is an optical brain imaging method that has become a valuable neuroimaging technique in research and clinical settings. The reliability of fNIRS hemodynamic measures is good to excellent (for review see Ref. 5). Test–retest performance has been examined in different conditions and age groups. Visual stimulation and finger-tapping measured over the occipital and motor cortex in healthy adults have been reported to have intraclass correlation coefficients (ICC) of 0.84 and 0.80, respectively. Rhythmic handgrip exercises measured over prefrontal lobe revealed similar reliability in healthy adults (ICC = 0.83) compared with patients with moderate to severe traumatic brain injury (ICC = 0.70). Functional brain networks during resting-state investigated using independent component analysis and graph metrics have been reported to exhibit fair (ICC = 0.56) to good (0.6 < ICC < 0.74) reliability, respectively. Excellent reliability (ICC = 0.89 to 0.94) has been shown in infants of ages 4 to 16 months during visual-auditory recordings. Studies of peripheral muscle in healthy adults during handgrip contractions found ICC = 0.86, and in patients with chronic heart failure during leg muscle cycling, ICC = 0.85 to 0.92.

In addition to the hemodynamic measures based on the two primary chromophores, i.e., oxygenated (HbO₂) and deoxygenated (HHb) hemoglobin (as assessed using wavelengths around 775 and 850 nm, respectively), fNIRS is also capable of monitoring a third chromophore, which is the mitochondrial cytochrome-c-oxidase (CCO), assessed using a wavelength around 810 nm (Fig. 1) (for review see Ref. 14). The optical signature of the CCO is predominantly the Cu₄ center, which is the binuclear copper site in soluble domains of CCO. fNIRS measures the oxidation state of mitochondrial CCO, termed oxCCO signal, which is an indicator of the CCO redox changes. CCO is the terminal electron acceptor in the mitochondrial respiratory chain and joint responsible for over 95% of the oxygen metabolism in the synthesis of adenosine triphosphate, the main source of cellular energy. The in vivo concentration of oxCCO is ~10% of that of hemoglobin, making it harder to detect. Yet, despite its small change in amplitude, the oxCCO signal is a remarkably stable signal, and has a higher brain-specificity compared with hemoglobin.

While the hemoglobin measures, HbO₂ and HHb, provide information on hemodynamic circulation and intravascular oxygenation, oxCCO is an indicator of cellular oxygen metabolism. Changes in oxCCO indicate an alteration in the balance between oxygen delivery and oxygen consumption. Hence, a more complete analysis of cerebral health requires concurrent monitoring of all three chromophores. Applications of this approach included assessment of neonatal brain injury or traumatic brain injury, bedside monitoring of cortical mitochondrial metabolic status in clinical settings, or the assessment of cognitive tasks, such as driving. These brain mitochondrial functions cannot be assessed using other neuroimaging methods, such as functional magnetic resonance imaging (fMRI).

Unlike the hemodynamic measures, the reliability of oxCCO measurements has so far not been reported. Reasons for this may be because it is harder to detect and requires special fNIRS instrumentation that is currently commercially not available.
In the present study, we therefore aimed to test–retest the oxCCO signal and to provide a first estimate of its reliability in a healthy population.

2 Materials and Methods

2.1 Participants

Ten physically and psychiatrically healthy adult volunteers [six females, mean age (±STD) = 34 ± 4.5 years] were recruited by advertising and posted flyers. Exclusion criteria were any psychiatric, or neurological disorder, unstable medical condition, or pregnancy. All participants gave written informed consent. The study was approved by the Institutional Review Board of the New York State Psychiatry Institute and performed in accordance with the Declaration of Helsinki.

2.2 Respiratory Challenges

Quantification of the oxCCO signal demands significant changes in the arterial oxygen saturation to be reliably detected in the cortex. Therefore, two respiratory challenges were chosen as previously described to systemically manipulate cerebral oxygenation via hypo- and hypercapnia, i.e., reduced and increased carbon dioxide in the blood. For the test–retest design, data from each participant were collected in two sessions at the same time of the day separated by one week. In each session, participants underwent both challenges.

**Hypocapnia** was induced by hyperventilation consisting of five repetitions of alternate periods of rapidly breathing in and out with constant respiratory volume (20 s) and normal breathing (40 s) with a total duration of five minutes.

**Hypercapnia** was induced by breath-holding consisting of five repetitions of alternate periods of breath hold (20 s) and normal breathing (40 s) with a total duration of five minutes. Participants were trained prior to recording to perform the inspirational breath-hold volume of air similar to a normal breath cycle to avoid inhaling larger volumes of air than the volume of a normal breath cycle.

Prior to and after each challenge, a resting-state of five minutes was collected, during which participants were asked to sit still with their eyes open to allow the hemodynamic cortical system to normalize. The resting-state periods were not included in the retest analysis.

2.3 Functional Near-Infrared Spectroscopy Instrumentation

An NIRO 300 instrument (Hamamatsu Photonics) was used to collect simultaneous measurements of the metabolic and hemodynamic fNIRS signals. The instrument delivers four wavelengths of light (775, 810, 850, and 910 nm, respectively, Fig. 1) by four light sources (pulsed laser diodes) that are detected by three closely placed detectors (photodiodes) (Fig. 2). The concentration changes of oxCCO, HbO₂, and HHb are measured by the middle photodiode, whereas TOI is measured by the use of all three. Illustration of the probe positioning of the NIRO 300 (Hamamatsu Photonics). Areas of the dorsolateral prefrontal cortex were identified using the Montreal Neurological Institute (MNI) space brain atlas converted to the International 10–20 System.
TOI can be approximated as the ratio of oxygenated to total tissue hemoglobin. The oxCCO, HbO2, and HHb concentration changes are measured by the middle photodiode, whereas TOI is measured by the use of all three. The light sources and detectors (∼10 × 10 mm²) were held in a light proof holder and set at a distance of 4.5 cm. Both light sources and detectors were fixed in a rubber probe to allow for direct skin contact. A fiberoptic plate was used as the detector window, allowing light to be conducted from the skin surface to the sensors without distorting the spatial distribution. Placed under a flexible head cap, two channels were positioned to cover parts of the dorsolateral prefrontal cortex (i.e., Fp1 and Fp2 according to the International 10–20 system33) (Fig. 2). Sampling rate was set to 1 Hz. Data were detrended, zero-meaned, and band-pass-filtered in the range 0.005 to 0.3 Hz, using a fifth-order Butterworth filter, to remove physiological noise, using functions of the HOMER2 processing package.34

Peripheral physiological measures were assessed using a capnometer (LifeSense LS1-9R, Nonin Medical) by means of the arterial tissue oxygen saturation (SpO2) and the partial end-tidal carbon dioxide (PetCO2) of the exhaled air. Participants wore a nasal cannula through which tidal gases were sampled at 1 Hz.

3 Data Analysis
All data analysis was performed using MATLAB® (Release 2017b, MathWorks, Massachusetts).

fNIRS data from both channels were averaged. Changes of the concentration of ΔoxCCO, ΔTOI, ΔHbO2, and ΔHHb were used as indicators of metabolic and hemodynamic activities.1,7,11 The five hypo- and hypercapnia intervals were block-averaged and relative concentration changes were computed from the end of the rest period (i.e., time-locked two seconds before onset of hypo-/hypercapnia) to the end of the challenge period (i.e., time-locked two seconds after offset of hypo-/hypercapnia). The subject-specific signal changes were then used for the reliability analysis.

Fig. 3 Measured signals. Experimentally measured peripheral physiological signals (PetCO2, SpO2) and fNIRS signals (oxCCO, TOI, HbO2, and HHb, baseline corrected) are indicated by black lines. Gray patches indicate the hypo- and hypercapnia intervals (20 s) alternating with rest periods of normal breathing (40 s). Data are shown for an example participant.
3.1 Reliability

As primary measure of test–retest reliability, the ICC was applied using the following previous recommendations:\(^5\)

- ICC was computed as type “3–1” according to the Shrout and Fleiss convention,\(^35\) i.e., a two-way mixed single measures.
- The two-way mixed model was chosen based on the assumption that the results may not be generalized to fNIRS instruments other than the NIRO 300 (since currently commercially not available). The type of single measures was chosen because only the NIRO 300 instrument was used.
- ICC was calculated with confidence intervals based on an alpha level of significance \( p < 0.05 \). The corresponding \( p \)-value of a hypothesis test with the null hypothesis \( ICC = 0 \) was reported.
- Following the guidelines for interpretation by Cicchetti,\(^36\) the ranges of ICC < 0.40, ICC 0.40 to 0.59, ICC 0.60 to 0.74, and ICC > 0.75, were considered to reflect poor, fair, good, and excellent test–retest reliability, respectively.

3.2 Agreement

As measure of agreement, Bland–Altman plots\(^37\)--\(^39\) were applied, which are based on the analysis of the differences between two measurements. Agreement quantifies how close two measurements made on the same subject are and is given in unit scale as the measurements themselves.\(^40\) The corresponding estimates of agreement, i.e., mean difference, with lower and upper 95% limits of agreement (LoA), help in interpretation of the ICC\(^35\) in that they uncover trends or consistent bias between measurement sessions.

3.3 Repeatability

Bland–Altman plots can also be used to assess the repeatability of a measure. The plot is then used to check whether the variability or precision is related to the size of the characteristic signal being measured. Optimally, for repeated measurements of the same instrument, the mean difference should be zero. Therefore, the coefficient of repeatability (CR) can be derived as two times the standard deviation of the differences between the two measurements.\(^37\)

As the CR is a measure in the same unit scale as the measurements themselves, the coefficient of variation (CV) was calculated as an additional measure of dispersion expressed in percentage (%). The CV is defined as the ratio of the standard deviation to the absolute value of the mean.

4 Results

The experimentally measured signals from an example participant are shown in Fig. 3. The peripheral physiological data
PetCO2 and SpO2, as well as the fNIRS data oxCCO, TOI, HbO2, and HHb, are illustrated over the time course of the five hypo- and hypercapnia intervals (gray patches). All signals showed a relative change in response to hypo- and hypercapnia and a return close to baseline values after each challenge.

The block-averaged signal changes “across sessions” reflected typical responses to hypo- and hypercapnia (Fig. 4). Both the time courses as well as the relative magnitudes of the oxCCO, ΔTOI, ΔHbO2, and ΔHHb signals obtained during the respiratory challenges are in agreement with previous work.19,28 In response to hypocapnia, we observed a decrease in ΔPetCO2 and an increase in ΔSpO2, whereas the fNIRS data showed a decrease in ΔoxCCO, ΔTOI, ΔHbO2, and an increase in ΔHHb. The opposite pattern was observed in response to hypercapnia, an increase in ΔPetCO2, a decrease in ΔSpO2, an increase in ΔoxCCO, ΔTOI, ΔHbO2, and a decrease in ΔHHb.

Importantly, there were no significant differences between sessions for any of the signals as assessed using ANOVA (one-way main effects of “session”: hypocapnia ΔPetCO2 $F = 1.102, \ p = 0.307$; hypercapnia ΔPetCO2 $F = 0.798, \ p = 0.383$; hypocapnia ΔSpO2 $F = 2.037, \ p = 0.171$; hypercapnia ΔSpO2 $F = 0.466, \ p = 0.503$; hypocapnia ΔoxCCO $F = 0.418, \ p = 0.526$; hypercapnia ΔoxCCO $F = 0.967$; hypocapnia ΔTOI $F = 1.346, \ p = 0.261$; hypercapnia ΔTOI $F = 0.018, \ p = 0.893$; hypocapnia ΔHbO2 $F = 3.366, \ p = 0.083$; hypercapnia ΔHbO2 $F = 1.367, \ p = 0.257$; hypocapnia ΔHHb $F = 0.657, \ p = 0.428$; hypercapnia ΔHHb $F = 1.036, \ p = 0.332$). The nonsignificance between sessions is considered to be an important aspect for ICC selection as it helps in determining the selection between one- or two-way models.5

The subject-specific signal changes were then used to estimate the reliability between sessions (Figs. 5 and 6, Table 1).
ICCs demonstrated good to excellent reliability for the peripheral measures $\Delta$PetCO$_2$ (hypo-/hypercapnia ICC = 0.518/745) and $\Delta$SpO$_2$ (ICC = 0.871/937). The NIRS data exhibited excellent reliability in all signals $\Delta$oxCCO (ICC = 0.933/876), $\Delta$TOI (ICC = 0.735/922), $\Delta$HbO$_2$ (ICC = 0.930/905), and $\Delta$HHb (ICC = 0.916/579) as illustrated in the scatter plots.

In accordance, Bland–Altman plots demonstrated good agreement for all signals without significant differences in the mean differences as assessed using ANOVA. The evenly scattered measurement points without trends suggested that there was no consistent bias between sessions (Figs. 5 and 6). The mean differences were smaller for the oxCCO signal compared with the hemodynamic signals (Table 1) indicating that the metabolic measurements separated by a week were closer.

Repeatability refers to the variation in a measure under identical conditions over which the underlying signal can be considered to be constant. Under such conditions, variability may then be ascribed to the measurement process itself. The CR is a precision measure representing the unit scale value below which the absolute difference between two repeated measurements may be expected to lie with a probability of 95%. In our case, we observed a smaller CR for the oxCCO signal (CR = 0.248/0.157) compared with the hemodynamic signals (Table 1) most likely because its in vivo concentration is $\sim$10% that of hemoglobin. Therefore, the CRs of the metabolic and the hemodynamic signals may not be directly comparable.

The CV as another measure of dispersion, however, is expressed as a percentage, thus allowing for a direct comparison between the metabolic and the hemodynamic signals. The CV also showed a relatively small value for the oxCCO signal (CV = 4.432%/3.694%) compared with the hemodynamic signals (Table 1) suggesting that the degree of variation between sessions was smaller. Notably, the CV of the peripheral measures

**Fig. 6** Reliability, agreement, and repeatability TOI, HbO$_2$, and HHb. (a) ICCs ($\pm$CI) and (b) scatter plots illustrating the reliability between sessions. (c) Bland–Altman plots illustrating the agreement and repeatability between sessions as assessed by the mean differences. Data are shown for both hypocapnia (black) and hypercapnia (gray). See Table 1 for statistics.
(PetCO₂, SpO₂) cannot be compared with those of the fNIRS signals because of the different units.

### 5 Discussion

Cortical CCO measurements derived from fNIRS can yield crucial information about cerebral mitochondrial metabolism. As its first demonstration in 1977, recent developments in instrumentation and analytical methods used to resolve the oxCCO signal have led to several clinical applications of the oxCCO signal. However, it belongs to those early-generation instruments that used only four discrete wavelengths (775, 810, 850, and 910 nm, respectively) to assess both the metabolic and hemodynamic signals. These wavelengths are in relatively good agreement with the optimal wavelengths reported for CCO measures. Nonetheless, this technology has been superseded by systems using broadband spectroscopy covering a near-infrared wavelength resolution from ∼780 to 900 nm that is thought to obtain a more robust oxCCO signal. There are also novel signal algorithms based on diffusion approximation, which have been suggested as an alternative to the modified Beer–Lambert’s law used in the present work for the deconvolution of oxCCO. Such algorithms are thought to avoid the possibility of cross talk between the hemodynamic and metabolic changes that might occur when using the MBL algorithm. It is thus possible that some of the drawbacks of the NIRO 300 instrumentation, such as low wavelength resolution and possible cross talk, might have affected the reliability of the present data. The exact evaluation of the corresponding instrumental behavior would, however, go beyond the scope of the present work. Using

### Table 1 Reliability, agreement, and repeatability. ICCs and Bland–Altman measures, i.e., the MD with lower and upper LoA, the CR and the CV, between sessions for both the hypo- and hypercapnia challenge. See Figs. 5 and 6 for illustration.

<table>
<thead>
<tr>
<th></th>
<th>ΔPetCO₂</th>
<th>ΔSpO₂</th>
<th>ΔoxCCO</th>
<th>ΔTOI</th>
<th>ΔHbO₂</th>
<th>ΔHHb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypocapnia</td>
<td>Reliability</td>
<td>ICC</td>
<td>0.793</td>
<td>0.667</td>
<td>0.827</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upper 95% CI</td>
<td>0.944</td>
<td>0.905</td>
<td>0.954</td>
<td>0.874</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower 95% CI</td>
<td>0.366</td>
<td>0.109</td>
<td>0.449</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.002</td>
<td>0.012</td>
<td>0.001</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>Agreement</td>
<td>MD</td>
<td>-3.860</td>
<td>0.015</td>
<td>0.062</td>
<td>-0.709</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upper 95% LoA</td>
<td>-0.633</td>
<td>0.053</td>
<td>0.310</td>
<td>1.762</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower 95% LoA</td>
<td>-7.088</td>
<td>-0.023</td>
<td>-0.185</td>
<td>-3.180</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>CR</td>
<td>3.228</td>
<td>0.038</td>
<td>0.248</td>
<td>2.471</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CV %</td>
<td>0.814</td>
<td>3.455</td>
<td>4.432</td>
<td>1.033</td>
</tr>
<tr>
<td>Hypercapnia</td>
<td>Reliability</td>
<td>ICC</td>
<td>0.687</td>
<td>0.833</td>
<td>0.913</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upper 95% CI</td>
<td>0.912</td>
<td>0.956</td>
<td>0.978</td>
<td>0.848</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower 95% CI</td>
<td>0.146</td>
<td>0.463</td>
<td>0.690</td>
<td>-0.143</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>0.010</td>
<td>0.001</td>
<td>0.000</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>Agreement</td>
<td>MD</td>
<td>0.597</td>
<td>0.008</td>
<td>-0.004</td>
<td>-0.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upper 95% LoA</td>
<td>2.910</td>
<td>0.036</td>
<td>0.154</td>
<td>2.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower 95% LoA</td>
<td>-1.717</td>
<td>-0.021</td>
<td>-0.161</td>
<td>-2.173</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>CR</td>
<td>2.314</td>
<td>0.029</td>
<td>0.157</td>
<td>2.108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CV %</td>
<td>0.648</td>
<td>22.016</td>
<td>3.694</td>
<td>5.708</td>
</tr>
</tbody>
</table>
technology reliability of the metabolic signal may even improve. This consideration has been accounted for in the ICC calculation (i.e., by choosing the corresponding type two-way mixed model with single measures).

Regarding timing, the test–retest data were collected in two sessions separated by one week. It may be assumed that, within this time period, the intervening respiratory challenges did not change the ability to induce hypo- and hypercapnia. Participants were further asked to maintain their breathing patterns between sessions such as the strength or depth of breath during hyperventilation. Thus, test–retest reliability should have not been compromised by timing. This assumption was supported by the good reliability observed for the respiratory parameters (PetCO2 ICC = 0.793/0.687, SpO2 ICC = 0.667/0.833) reflecting consistency of the administered respiratory challenges (Figs. 5 and 6, Table 1). Longer time spans, in contrast, such as a year or more, may affect reliability or circumstances where individuals experience changes in pulmonary, metabolic, or vascular systems, reflecting a change in signal and its measurement.

Regarding probe positioning, the test–retest data were collected from the same cortical areas, repositioned by the same experimenter. The probe itself, however, consisted of only a small number of channels. Nonetheless, this aspect was considered of minor relevance for the purpose of test–retesting as the spatial distribution of oxCCO changes in human brain requires further investigation before being quantified.

In other words, we expected to deal with global (nonlocalized) changes across the whole prefrontal cortex suggested to be unchanged between sessions. Considering cerebral pathological changes, in contrast, would unlikely be expected to have a uniform distribution across the brain, such as in the case of known brain disease or brain injury and may then require regional (localized) measures of oxCCO. Signal alterations in such cases may be smaller than those tested in this study and may require measurements of greater reliability.

6 Conclusion
Taken together, we found robust reliability of both the metabolic oxCCO signal and the hemodynamic signals derived from fNIRS in the human brain. These findings suggest that the oxCCO signal can be reliably measured under conditions of respiratory challenges to be implemented effortlessly in both research and clinical settings.

Disclosures
The authors declare no conflicts of interest.

Acknowledgments
The work was funded by a Janssen Fellowship in Translational Neuroscience at Columbia University, New York, awarded to LH.

References


P. Phan et al., “A new multichannel broadband near infrared spectroscopy system to measure the spatial distribution of cytochrome-c-oxidase and tissue oxygenation,” in Biomedical Optics, OSA (2016).


Lisa Holper received her MD degree from the University Medicine Berlin and has done her residency at the Department of Psychiatry, Psychotherapy, and Psychosomatics, Hospital of Psychiatry Zurich, University Zurich. Currently, she is working as a research fellow at the Columbia University. She has published and contributed to over 30 papers addressing human research in healthy and psychiatric populations particularly in the field of optical brain imaging.

J. John Mann is the Paul Janssen professor of translational neuroscience and vice chair for research in the Department of Psychiatry at Columbia University. His research employs functional brain imaging, neurochemistry, and molecular genetics to probe the causes of depression and suicide. He has published 458 papers and edited 10 books on the subjects of the biology and treatment of mood disorders, suicidal behavior, and other psychiatric disorders.