

# The reliability of $^{31}\text{P}$ -MRS and NIRS measurements of spinal muscle function

Jonathan Fulford<sup>1</sup>, Agris Liepa<sup>2</sup>, Alan R. Barker<sup>2</sup> & Judith R. Meakin<sup>3</sup>

<sup>1</sup>Exeter NIHR Clinical Research Facility, University of Exeter Medical School, University of Exeter, UK

<sup>2</sup>Sport and Health Sciences, College of Life and Environmental Sciences, University of Exeter, UK

<sup>3</sup>Biomedical Physics, College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK

## *Corresponding author:*

Dr Judith R. Meakin, Physics Building, University of Exeter, Stocker Road, Exeter, EX4 4QL, UK

Email: [j.r.meakin@exeter.ac.uk](mailto:j.r.meakin@exeter.ac.uk)

Tel: +44 (0)1392 724109

Fax: +44 (0)1392 724111

## **ABSTRACT**

Phosphorous magnetic resonance spectroscopy ( $^{31}\text{P}$ -MRS) and near-infra red spectroscopy (NIRS) provide methods for measuring spinal muscle function non-invasively but their reliability is not established. The aim of this study was assess the reliability (ICC) and error magnitude (CV%) of measurements of muscle phosphocreatine (PCr), tissue oxygenation index (TOI), and muscle deoxyhaemoglobin (HHb) acquired during fatigue and in recovery after 24 s exercise in the lumbar muscles. Ten healthy participants (19-25 years, 5 male, 5 female) performed exercise that involved holding the upper body unsupported in slight extension until fatigue and then, after 30 minutes of rest, for repeated bursts of 24 seconds. ICCs indicated good to excellent reliability of baseline measures (TOI:0.75) and of amplitude changes during fatigue (PCr:0.73, TOI:0.69, HHb:0.80), and recovery (HHb:0.96) and poor to fair reliability for time constants describing rates of change during fatigue (PCr:0.11) and recovery (PCr:0.31, HHb:0.47). CV% indicated varying relative measurement error across baseline measures (TOI:5%), amplitude changes during fatigue (PCr:7%, TOI:38%, HHb:31%) and recovery (HHb:31%), and in time constants for fatigue (PCr:39%) and recovery (PCr:20%, HHb:37%). The results suggested that reliability would be sufficient for future studies on spinal muscle function but that measurement error may be too large to evaluate individuals.

## **KEYWORDS:**

Lumbar spine muscles; Phosphocreatine; Tissue oxygenation index; Deoxyhaemoglobin

## INTRODUCTION

The lumbar extensor muscles, particularly multifidus and erector spinae, play an important role in providing mechanical stability and controlling movement of the lumbar spine and trunk [14].

Determining how these muscles function in vivo is important for understanding spinal biomechanics and for developing appropriate therapeutic treatments to address its dysfunction.

Spinal muscle function can be explored using a wide variety of methods [8]. Phosphorous magnetic resonance spectroscopy ( $^{31}\text{P}$ -MRS) and near-infra-red spectroscopy (NIRS) provide ways of determining muscle function and are advantageous over other methods (e.g. tissue biopsy) because they are non-invasive.  $^{31}\text{P}$ -MRS can be used to quantitatively monitor intramuscular metabolites such as phosphocreatine (PCr), of which the rate of depletion during exercise, and the rate of recovery after exercise, provides information on muscle metabolism and oxidative capacity [28]. NIRS can be used to assess a range of blood characteristics, such as the proportion of the blood within the muscle tissue that is oxygenated and modifications in the amount of deoxygenated blood. Such measures provide an indication of the availability of oxygen for metabolism and provide a marker of oxygen extraction [29].

$^{31}\text{P}$ -MRS and NIRS have been used to determine muscle function in many parts of the body including the lumbar spine, albeit to a lesser extent. They have the potential, particularly if used in combination, to provide useful insights in a range of situations where lumbar extensor muscle strength and endurance is impaired such as in low back pain patients or elderly individuals at risk of falling, and to help evaluate the effectiveness of exercise therapies. The methods are generally thought to be reliable [10]; however, reliability has not been assessed in the spinal muscles.

These pose additional technical challenges due to their relatively small size and close proximity

to bone and so the aim of this study was to assess the reliability of  $^{31}\text{P}$ -MRS and NIRS measurements for assessing the function of the lumbar extensor muscles during exercise and recovery and to characterise the errors in these measurements.

## **METHODS**

### **Participants**

Ten healthy volunteers, with no lower back pain or known musculoskeletal disease, were recruited via convenience sampling and gave their informed consent to participate. The participants (five male and five female), were aged 19 to 25 years with a mean body mass of 68 kg (sd = 12 kg) and a mean height of 173 cm (sd = 11 cm). Each participant attended three study visits with at least seven days between each visit. Approval for the study was gained from a local research ethics committee and the study met the ethical standards of the journal [15].

### **Exercise protocol**

At each visit the participants performed an exercise protocol within the bore of a 1.5 T superconducting magnet (Intera, Philips, The Netherlands). Participants were initially positioned in a straight-legged supine posture and imaging was performed to obtain anatomical information. The participants were then positioned prone using a set-up similar to that of Rzanny et al. [30], with padding under the hips, straps securing their legs to the scanner bed, and a foam wedge under their upper body (Figure 1a). A NIRS probe was attached to the skin over the left hand side muscle bulk (Figure 1b) at the level of the L3/L4 intervertebral disc (location estimated by palpation). A 6 cm  $^{31}\text{P}$  coil was positioned immediately above the NIRS probe using a custom holder (Figure 1c).

The foam wedge was removed and participants asked to maintain the position of their upper body until fatigued (Figure 1d), with the endurance time recorded to the nearest second. NIRS and  $^{31}\text{P}$ -MRS data were recorded throughout. Following exhaustion, the participants were returned to the supine position to rest for a period of approximately 30 minutes. Participants were then repositioned prone, as detailed above, and asked to perform the same exercise as for the fatigue protocol but this time for only 24 s at which point the foam wedge was replaced and the participant relaxed for 216 s. This non-fatiguing, intermittent protocol was repeated four times by each participant.  $^{31}\text{P}$ -MRS and NIRS data were recorded throughout exercise and recovery.

### **MR spectroscopy**

$^{31}\text{P}$ -MRS data were acquired during exercise and recovery every 1.5 s with a spectral width of 1,500 Hz and 1,000 data points. Phase cycling with four phase cycles was employed, leading to a spectra being acquired every 6 s. The acquired spectra were quantified via peak fitting, using the jMRUI (version 3) software package employing the AMARES fitting algorithm [36]. Spectra were fitted assuming the presence of the following peaks:  $\text{P}_i$ , phosphodiester, Accepted version PCr,  $\alpha$ -ATP (2 peaks, amplitude ratio 1:1),  $\gamma$ -ATP (2 peaks, amplitude ratio 1:1), and  $\beta$ -ATP (3 peaks, amplitude ratio 1:2:1). PCr depletion and recovery responses were fitted with Prism 5 software (GraphPad Software Inc, La Jolla, California) by a single exponential.

For depletion (during fatigue exercise):

$$\text{PCr}(t) = \text{PCr}_{100} - \Delta\text{PCr}_{\text{ex}} \left(1 - e^{-t/\tau\text{PCr}_{\text{ex}}}\right) \quad \text{Equation 1}$$

where  $\text{PCr}_{100}$  is the value of PCr at the start of exercise (defined as 100%),  $\Delta\text{PCr}_{\text{ex}}$  is the difference between the PCr at the beginning of exercise and the exponential plateau,  $t$  is the time from the beginning of exercise, and  $\tau\text{PCr}_{\text{ex}}$  is the time constant for the exponential decay of PCr.

For recovery (after 24 s exercises):

$$\text{PCr}(t) = \text{PCr}_{\text{end}} + \Delta\text{PCr}_{\text{rec}}(1 - e^{-t/\tau\text{PCr}_{\text{rec}}}) \quad \text{Equation 2}$$

where  $\text{PCr}_{\text{end}}$  is the value at the end of exercise,  $\Delta\text{PCr}_{\text{rec}}$  is the difference between the PCr at end exercise and fully recovered,  $t$  is the time from exercise cessation, and  $\tau\text{PCr}_{\text{rec}}$  is the time constant for the exponential recovery of PCr. Each of the four 24 s recovery periods were fitted individually and derived values of  $\tau\text{PCr}_{\text{rec}}$  averaged.

### **Near-infra red spectroscopy**

The intensity of the NIRS incident and transmitted light was recorded continuously at 1 Hz during the fatigue and 24 s exercise protocols (NIRO200, Hamamatsu Photonics KK, Japan) and the data used to determine the pre-exercise tissue oxygenation index (TOI) and the change in tissue oxygenation ( $\Delta\text{TOI}$ ) and deoxyhaemoglobin ( $\Delta\text{HHb}_{\text{ex}}$ ) during fatigue exercise.

The time course of the deoxyhaemoglobin signal during recovery after the 24 s exercise was fitted to a single exponential (within Prism) of the form:

$$\text{HHb}(t) = \text{HHb}_{100} - \Delta\text{HHb}_{\text{rec}}(1 - e^{-t/\tau\text{HHb}_{\text{rec}}}) \quad \text{Equation 3}$$

where  $\text{HHb}_{100}$  is the value at the end of exercise,  $\Delta\text{HHb}_{\text{rec}}$  is the difference between the values at end exercise and at the exponential plateau,  $t$  is the time from exercise cessation, and  $\tau\text{HHb}_{\text{rec}}$  is the time constant for the exponential decay. Each 24 s recovery period was fitted individually and the time constants determined for each, before being averaged.

## Statistics

Statistical analyses were performed using PASW 18.0 (SPSS Inc., Chicago, USA) with a 5 % level of significance. Repeated measures analysis of variance was used to test for within-subject effects and linear contrasts across the three visits; the Greenhouse-Geisser correction was used on data that did not meet the assumption of sphericity. Reliability was assessed by calculating single and average measures intra-class correlation coefficients (ICC) using a one-way random model. ICCs were classed as being poor ( $0 < \text{ICC} < 0.4$ ), fair ( $0.4 < \text{ICC} < 0.59$ ), good ( $0.60 < \text{ICC} < 0.74$ ), or excellent ( $0.75 < \text{ICC} < 1$ ) [4]. Measurement error was assessed by calculating the within-subject standard deviation ( $s_w$ ) and the within-subject coefficient of variation (CV%). Pearson correlation coefficients between the participants' standard deviation and mean measurements were calculated to determine whether the measurement error was proportional or fixed. Sample sizes required for future studies concerned with detecting longitudinal changes in measurements of an intervention group compared a control group were calculated for a range of effect sizes using:

$$N = \frac{8(Z_{\alpha} + Z_{\beta})^2(1-\rho)}{\varepsilon^2} \quad \text{Equation 4}$$

Where N is the total sample size,  $Z_{\alpha} = 1.96$  (significance level,  $\alpha = 5\%$ ),  $Z_{\beta} = 0.84$  (power,  $\beta = 80\%$ ),  $\rho$  is the correlation between the measurements taken at baseline and follow-up (assumed to be equal to the single-measures ICC determined in this study), and  $\varepsilon$  is the effect size (the anticipated mean change in the variable from baseline to follow-up divided by the sample standard deviation at baseline) [12].

## RESULTS

All ten participants completed three visits; the mean (range) interval between visits was 28 (8 to 49) days for visits 1 and 2, 13 (7 to 25) days for visits 2 and 3, and 40 (19 to 66) days for visits 1 and 3. The mean value of each  $^{31}\text{P}$ -MRS and NIRS measure is given in Table 1 by visit and as a pooled value. The within subject-effects and linear contrasts are also given in Table 1.

Participants held their trunk unsupported for a mean endurance time of 67 s (range 37 to 150 s) and during this exercise, PCr and TOI decreased ( $\Delta\text{TOI}$  and  $\Delta\text{PCr}_{\text{ex}}$ ) whilst HHb increased ( $\Delta\text{HHb}_{\text{ex}}$ ). In the recovery period after 24 s of exercise, HHb levels reduced ( $\Delta\text{HHb}_{\text{rec}}$ ) and PCr increased. With the exception of TOI, there were no significant within-subject effects or linear contrasts (Table 1).

The single measures reliability (Table 2) of the measurements of TOI and changes in TOI ( $\Delta\text{TOI}$ ), PCr ( $\Delta\text{PCr}_{\text{ex}}$ ), and HHb ( $\Delta\text{HHb}_{\text{ex}}$  and  $\Delta\text{HHb}_{\text{rec}}$ ) ranged from good to excellent. The time constants ( $\tau\text{PCr}_{\text{ex}}$ ,  $\tau\text{PCr}_{\text{rec}}$ ,  $\tau\text{HHb}_{\text{rec}}$ ) characterising these changes, however, had poor to fair reliability. The average measures ICCs (Table 2) showed that reliability would be improved if multiple measurements were taken.

The within-subject standard deviations (Table 2) demonstrate the magnitude of the measurement error for each  $^{31}\text{P}$ -MRS and NIRS measure. For two of the measures ( $\tau\text{PCr}_{\text{ex}}$  and  $\Delta\text{TOI}$ ), there was a significant correlation (Table 2) between the subject standard deviation and mean, suggesting that the error in these measures was proportional rather than fixed. The within-subject CV (Table 2), indicating the relative error on each measure, varied from 5 % to 39 %. Figure 2 shows a scatter plot of the single measures ICC versus coefficient of variation to allow visual



comparison of how the eight measures performed in terms of reliability and measurement error. The estimated sample sizes required to investigate longitudinal changes in the  $^{31}\text{P}$ -MRS and NIRS measures are shown in Table 3 for three different effect sizes.

## **DISCUSSION**

The aim of this study was to assess the reliability and measurement error in  $^{31}\text{P}$ -MRS and NIRS measurements of lumbar muscle function after exercise and recovery. The exercise protocol used in this study was similar to the Biering-Sorensen test that is commonly used to assess the endurance of spinal extensor muscles in the clinical and research setting [7] and involves subjects maintaining their unsupported trunk in a horizontal supine position for as long as possible. To allow the test to be performed within the confines of the scanner it was modified slightly by raising the hips and including a small amount of extension. This produced a set-up similar to the Ito test [26] (a modification of the Biering-Sorensen test that involves less contribution from the hip muscles [26]) and is similar to that used in previous studies on  $^{31}\text{P}$ -MRS of the extensor muscles [30]. Isometric exercise in slight extension has been shown to be effective at inducing fatigue in the back extensors [5] with a significant reduction in strength (i.e. fatigue) occurring after subjects performed a 45 second hold at 10 degrees of extension [5]. The mean endurance time in the current study was lower than in many others [7] reflecting the higher levels of muscle activity induced by the slight trunk extension [25] and the extended position of the arms. The reliability and variation in the endurance time was consistent with other studies [7, 18, 21], and the lack of significant differences or linear trends (Table 1) suggested that no learning effects occurred.

In accordance with guidelines for reporting reliability, [20] intra-class correlation coefficients (ICCs) were calculated to assess the reliability of the  $^{31}\text{P}$ -MRS and NIRS measures, and within-subject standard deviations,  $s_w$ , and coefficients of variation, CV%, were calculated to assess the measurement errors. ICC indicates the ability of a method to determine differences between participants despite the presence of measurement error and, for this study, provided a useful way of comparing the performance of the eight  $^{31}\text{P}$ -MRS and NIRS measures. The inherent dependence of the ICC on the sample heterogeneity, however, means that the ICC values might not be generalizable to all other samples or populations [6]. The  $s_w$  provides useful information for future studies that intend to measure changes within subjects as, when multiplied by 2.77, it can be used to estimate the smallest measurable difference (SMD) [2]. This is the value below which 95% of repeated measurements will lie if there is no difference between them and thus indicates the value above which a measured difference can be accepted as being a true difference [2]. For errors that are proportional to the measurement size (as found for  $\tau\text{PCr}_{\text{ex}}$  and  $\Delta\text{TOI}$  in this study) the CV% provides a more meaningful statistic and can also be used to estimate the smallest measurable difference [2]. As the CV% indicates the relative error, it also provided a useful way of comparing the errors across the eight  $^{31}\text{P}$ -MRS and NIRS measures in this study. Similar to the ICC, however, the CV% values might not be generalizable to all other samples and populations.

$^{31}\text{P}$ -MRS measurements have been used in the lumbar spine to monitor PCr during fatigue [30], to explore muscle tension [33], and evaluate a short-term exercise intervention [17]. The reliability of  $^{31}\text{P}$ -MRS measurements has been assessed in the thigh [22] and calf [10] and found to be similar to that determined in the current study. A potential limitation when undertaking PCr measurements in the spine, compared to other regions that are commonly assessed using  $^{31}\text{P}$ -

MRS, such as the legs, is the inherent low SNR that reduces the confidence of PCr intensity determinations at each time point and thus the subsequent exponential fits. This low SNR is principally a result of the small muscle masses; however, an additional problem in the spine is the proximity of the large amount of bone (i.e. vertebrae) that causes magnetic field homogeneity to be poor relative to equivalent measurements on regions such as the legs. This low SNR is reflected in the low reliability and large measurement error in the estimate of the time constants for PCr depletion; the reliability of the relative depletion of PCr at the point of fatigue utilizing the same data, however, was excellent with relatively small measurement error.

NIRS has previously been used to understand more about the mechanisms underlying fatigue in the lumbar spine muscles [19, 23, 37] and to assess exercise therapy [27]. NIRS measures are generally considered reliable [29] and previous studies on the lumbar extensor muscles have found the reliability of changes in response to exercise to be fair for oxygen saturation [9], good for muscle oxygenation [18], and excellent for blood volume [18]. As a result of NIRS being a surface based methodology, where sampling takes place for a limited depth beneath the area the emission/detection probes are placed, it is extremely sensitive to the exact location of placement. Thus, in addition to variations in placement leading to different muscle regions being sampled, small variations in, for example, subcutaneous fat thicknesses can have significant impact on the amplitude of HHb changes [11]. We are not aware of previous studies investigating the reliability of recovery kinetics in the lumbar spine but a study on recovery in the gastrocnemius [3] demonstrated that the magnitude of the change in HHb was more reliable than the time constant for the rate of change. A similar finding was found in the present study, with the amplitude of change during recovery having better reliability and slightly lower measurement error than the time constant.

The results of this study have implications for future investigations on the function of the lumbar extensor muscles in vivo. A reduction in the strength and endurance of the lumbar extensor muscles has been implicated as a contributory factor to both low back pain (LBP) [35] and risk of falling in the elderly [13]. LBP is a major health problem, with a lifetime prevalence of around 85 %. It also represents the most common cause of work related disability in people under the age of forty five [31] and incurs high societal and economic costs [24]. Age related muscle atrophy affects around 50 % of those over the age of 60 leading to functional impairment and disability [16] and, in the trunk, increases the risk of falling [13]. As a result, work has been undertaken to assess lumbar muscle atrophy in LBP patients and the elderly, leading to the development of conditioning programmes aimed at improving the strength, endurance and neuromuscular control in the muscles around the spine [32]. To examine the effectiveness of any intervention methodology, it is important to have techniques that can reliably evaluate the response of an individual or a group. In terms of evaluating individuals, longitudinal changes need to be larger than the SMD and for the assessment of groups, methods should be reliable enough for studies to be performed with a realistic number of subjects [1]. Using the reliability results from the current study we have predicted that the sample sizes required to detect a medium (0.5) sized effect in an intervention group compared to a control group range from 10 to 225 (Table 3).

The novelty of acquiring  $^{31}\text{P}$ -MRS and NIRS measures in the spine muscles means that there is little evidence for the size of the response that would be clinically relevant for an exercise intervention in LBP patients and the elderly. Pilot data from healthy individuals undergoing a 4-week exercise intervention showed  $\Delta\text{PCr}_{\text{ex}}$  changes that were comparable to the SMD difference but produced an effect size of 0.5 [17]. Data from LBP patients undergoing a 4-week

intervention suggests that changes in NIRS amplitudes and time constants may be too small to detect in individuals and that the effect size may be greater than 0.5 for amplitude changes but less than 0.5 for time constants [27]. In both of these studies, however, the intervention was considerably shorter than that typically used in clinical intervention studies [32, 34] and therefore likely to be producing a smaller response than is clinically relevant. Other observational studies comparing healthy volunteers and LBP patients [19, 27] suggest similar results for NIRS amplitude changes and time constants, whereas an experimental study investigating spinal muscles in different postures shows that differences in TOI can be much greater than the SMD [9]. This evidence, although sparse, indicates that the measures investigated in this study may not be appropriate for evaluating individuals but that it would probably be feasible to use most of the measures for evaluating groups. The relatively large errors and low reliability of the time constants, however, indicates that studies investigating these measures may require sample sizes of several hundred participants. This may prove prohibitive due to the practicalities and costs of recruiting and scanning participants but further studies are required to provide estimates of the effect sizes for clinically relevant exercise interventions.

## **ACKNOWLEDGMENTS**

We thank the participants who volunteered to take part in our study. We also thank the NIHR Clinical Research Facility, University of Exeter for funding JF's salary and the University of Exeter's Bridging the Gaps project (funded by the EPSRC's Bridging the Gaps initiative (grant reference EP/I001433/1)) for funding the project.

## REFERENCE LIST

1. *Atkinson G, Nevill AM*. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998; 26: 217-238
2. *Bland M*. An Introduction to Medical Statistics. 3rd ed. Oxford: Oxford University Press, 2000
3. *Buchheit M, Ufland P, Haydar B, Laursen PB, Ahmaidi S*. Reproducibility and sensitivity of muscle reoxygenation and oxygen uptake recovery kinetics following running exercise in the field. *Clin Physiol Funct Imaging* 2011; 31: 337-346
4. *Cicchetti DV*. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994; 6: 284-290
5. *Corin G, Strutton PH, McGregor AH*. Establishment of a protocol to test fatigue of the trunk muscles. *Br J Sports Med* 2005; 39: 731-735
6. *de Vet HCW, Terwee CB, Knol DL, Bouter LM*. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006; 59: 1033-1039
7. *Demoulin C, Vanderthommen M, Duysens C, Crielaard JM*. Spinal muscle evaluation using the Sorensen test: A critical appraisal of the literature. *Joint Bone Spine* 2006; 73: 43-50
8. *Demoulin C, Crielaard J-M, Vanderthommen M*. Spinal muscle evaluation in healthy individuals and low-back-pain patients: a literature review. *Joint Bone Spine* 2007; 74: 9-13
9. *Dupeyron A, Lecocq J, Vautravers P, Pélissier J, Perrey S*. Muscle oxygenation and intramuscular pressure related to posture and load in back muscles. *Spine J* 2009; 9: 754-759

10. *Edwards LM, Tyler DJ, Kemp GJ, Dwyer RM, Johnson A, Holloway CJ, Nevill AM, Clarke K.* The reproducibility of 31-phosphorus MRS measures of muscle energetics at 3 Tesla in trained men. *PLoS One* 2012; 7: e37237
11. *Feng W, Haishu D, Fenghua T, Jun Z, Qing X, Xianwu T.* Influence of overlying tissue and probe geometry on the sensitivity of a near-infrared tissue oximeter. *Physiol Meas* 2001; 22: 201-208
12. *Friedman LM, Furberg CD, DeMets DL.* Fundamentals of Clinical Trials. 4th ed. New York: Springer, 2010
13. *Granacher U, Gollhofer A, Hortobágyi T, Kressig RW, Muehlbauer T.* The importance of trunk muscle strength for balance, functional performance, and fall prevention in seniors: A systematic review. *Sports Med* 2013; 43: 627-641
14. *Hansen L, de Zee M, Rasmussen J, Andersen TB, Wong C, Simonsen EB.* Anatomy and biomechanics of the back muscles in the lumbar spine with reference to biomechanical modeling. *Spine* 2006; 31: 1888-1899
15. *Harriss DJ, Atkinson G.* Ethical standards in sport and exercise science research: 2014 update. *Int J Sports Med* 2013; 34: 1025-1028
16. *Janssen I, Heymsfield SB, Ross R.* Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. *J Am Geriatr Soc* 2002; 50: 889-896

17. *Juroskova V, Fulford J, Barker AR, Meakin JR*. Measuring spinal muscle metabolism in vivo. *Bone Joint J* 2014; 96-B (Supp 4): 16
18. *Kell RT, Farag M, Bhambhani Y*. Reliability of erector spinae oxygenation and blood volume responses using near-infrared spectroscopy in healthy males. *Eur J Appl Physiol* 2004; 91: 499-507
19. *Kell RT, Bhambhani Y*. Relationship between erector spinae static endurance and muscle oxygenation-blood volume changes in healthy and low back pain subjects. *Eur J Appl Physiol* 2006; 96: 241-248
20. *Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL*. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64: 96-106
21. *Latimer J, Maher CG, Refshauge K, Colaco I*. The Reliability and Validity of the Biering–Sorensen Test in Asymptomatic Subjects and Subjects Reporting Current or Previous Nonspecific Low Back Pain. *Spine* 1999; 24: 2085
22. *Layec G, Bringard A, Le Fur Y, Vilmen C, Micallef JP, Perrey S, Cozzone PJ, Bendahan D*. Reproducibility assessment of metabolic variables characterizing muscle energetics in vivo: A 31P-MRS study. *Magn Reson Med* 2009; 62: 840-854
23. *Maikala RV*. Medical utility of near infrared spectroscopy in low back musculoskeletal health and disorders. *Journal of Near Infrared Spectroscopy* 2012; 20: 151-184



24. *Mathew J, Singh SB, Garis S, Diwan AD*. Backing up the stories: the psychological and social costs of chronic low-back pain. *International Journal of Spine Surgery* 2013; 7: e29-e38
25. *Mayer JM, Verna JL, Manini TM, Mooney V, Graves JE*. Electromyographic activity of the trunk extensor muscles: Effect of varying hip position and lumbar posture during Roman chair exercise. *Arch Phys Med Rehabil* 2002; 83: 1543-1546
26. *Müller R, Strässle K, Wirth B*. Isometric back muscle endurance: an EMG study on the criterion validity of the Ito test. *J Electromyogr Kinesiol* 2010; 20: 845-850
27. *Olivier N, Thevenon A, Berthoin S, Prieur F*. An exercise therapy program can increase oxygenation and blood volume of the erector spinae muscle during exercise in chronic low back pain patients. *Arch Phys Med Rehabil* 2013; 94: 536-542
28. *Paganini AT, Foley JM, Meyer RA*. Linear dependence of muscle phosphocreatine kinetics on oxidative capacity. *Am J Physiol Cell Physiol* 1997; 272: C501-C510
29. *Pereira MIR, Gomes PSC, Bhambhani YN*. A brief review of the use of near infrared spectroscopy with particular interest in resistance exercise. *Sports Med* 2007; 37: 615-624
30. *Rzanny R, Grassme R, Reichenbach JR, Rottenbach M, Petrovitch A, Kaiser WA, Scholle HC*. Simultaneous surface electromyography (SEMG) and <sup>31</sup>P-MR spectroscopy measurements of the lumbar back muscle during isometric exercise. *J Neurosci Methods* 2004; 133: 143-152
31. *Sehgal N, Fortin JD*. Internal disc disruption and low back pain. *Pain Physician* 2000; 3: 143-157

32. *Standaert CJ, Herring SA*. Expert opinion and controversies in musculoskeletal and sports medicine: core stabilization as a treatment for low back pain. *Arch Phys Med Rehabil* 2007; 88: 1734-1736
33. *Strobel ES, Krapf M, Suckfüll M, Brückle W, Fleckenstein W, Müller W*. Tissue oxygen measurement and <sup>31</sup>P magnetic resonance spectroscopy in patients with muscle tension and fibromyalgia. *Rheumatol Int* 1997; 16: 175-180
34. *Suri P, Kiely DK, Leveille SG, Frontera WR, Bean JF*. Increased trunk extension endurance is associated with meaningful improvement in balance among older adults with mobility problems. *Arch Phys Med Rehabil* 2011; 92: 1038-1043
35. *Taimela S, Negrini S, Paroli C*. Functional rehabilitation of low back disorders. *Eura Medicophys* 2004; 40: 29-36
36. *Vanhamme L, van den Boogaart A, Van Huffel S*. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *J Magn Reson* 1997; 129: 35-43
37. *Yoshitake Y, Ue H, Miyazaki M, Moritani T*. Assessment of lower-back muscle fatigue using electromyography, mechanomyography, and near-infrared spectroscopy. *Eur J Appl Physiol* 2001; 84: 174-179

Table 1.  $^{31}\text{P}$ -MRS and NIRS measurements.

	Visit			Pooled	Within subject	Within subject	
	1	2	3		effects	linear contrasts	
					F(P)	F(P)	
During fatigue exercise	Endurance (s)	72 (51 to 93)	64 (43 to 85)	65 (49 to 81)	67 (48 to 86)	2.03 (0.16)	2.66 (0.14)
	$\Delta\text{PCr}_{\text{ex}}$ (%)	-71 (-77 to -65)	-71 (-78 to -64)	-70 (-78 to -62)	-71 (-77 to -64)	0.15 (0.86)	0.12 (0.73)
	$\tau\text{PCr}_{\text{ex}}$ (s)	15 (8 to 22)	15 (10 to 20)	18 (13 to 23)	16 (13 to 20)	0.54 (0.59)	1.23 (0.30)
	TOI (%)	79 (74 to 84)	81 (76 to 86)	83 (76 to 91)	81 (75 to 87)	2.93 (0.08)	5.37 (0.05)
	$\Delta\text{TOI}$ (%)	-16 (-24 to -8)	-14 (-22 to -6)	-13 (-18 to -7)	-14 (-21 to -8)	0.90 (0.42)	2.11 (0.18)
	$\Delta\text{HHb}_{\text{ex}}$ (AU)	10 (5 to 15)	9 (4 to 13)	8 (3 to 12)	9 (4 to 13)	1.30 (0.30)	2.03 (0.19)
After 24 s exercise	$\tau\text{PCr}_{\text{rec}}$ (s)	23 (18 to 28)	26 (22 to 30)	26 (22 to 30)	25 (22 to 28)	1.10 (0.35)	1.05 (0.33)
	$\Delta\text{HHb}_{\text{rec}}$ (AU)	-6 (-8 to -3)	-7 (-11 to -2)	-5 (-8 to -2)	-5 (-8 to -2)	1.50 (0.26)	1.90 (0.22)
	$\tau\text{HHb}_{\text{rec}}$ (s)	14 (8 to 21)	17 (11 to 23)	14 (10 to 19)	15 (10 to 19)	0.11 (0.90)	0.28 (0.62)

Mean (95% CI) values given by visit and pooled over the three visits, together with the within-subject effects and linear contrasts (F-statistic and P-value).

Table 2. Reliability and error statistics.

		Correlation between subject standard deviation and mean	Within-subject standard deviation, $s_w$	Within-subject coefficient of variation, CV (%)	Intraclass correlation coefficient, ICC	
		R (P)			Single measures	Average measures
During fatigue exercise	Endurance (s)	0.26 (0.48)	10 (7 to 14)	17 (9 to 22)	0.86 (0.66 to 0.96)	0.95 (0.85 to 0.99)
	$\Delta PC_{\text{ex}}$ (%)	-0.10 (0.78)	5 (3 to 7)	7 (4 to 9)	0.73 (0.43 to 0.92)	0.89 (0.69 to 0.97)
	$\tau PC_{\text{ex}}$ (s)	0.74 (0.01)	7 (5 to 10)	39 (23 to 51)	0.11 (-0.21 to 0.57)	0.26 (-1.09 to 0.80)
	TOI (%)	0.45 (0.19)	4 (3 to 6)	5 (2 to 7)	0.75 (0.46 to 0.92)	0.90 (0.72 to 0.97)
	$\Delta TOI$ (%)	0.65 (0.04)	6 (4 to 8)	38 (23 to 49)	0.69 (0.36 to 0.90)	0.87 (0.62 to 0.96)
	$\Delta HHb_{\text{ex}}$ (AU)	0.45 (0.19)	3 (2 to 4)	31 (19 to 40)	0.80 (0.55 to 0.94)	0.92 (0.78 to 0.98)
After 24 s exercise	$\tau PC_{\text{rec}}$ (s)	0.40 (0.25)	5 (3 to 7)	20 (11 to 26)	0.31 (-0.06 to 0.72)	0.58 (-0.21 to 0.88)
	$\Delta HHb_{\text{rec}}$ (AU)	-0.53 (0.14)	1 (1 to 1)	31 (11 to 43)	0.96 (0.87 to 0.99)	0.99 (0.95 to 1.00)
	$\tau HHb_{\text{rec}}$ (s)	0.29 (0.44)	5 (3 to 7)	37 (17 to 50)	0.47 (0.01 to 0.86)	0.72 (0.03 to 0.95)

Correlation is given as the Pearson correlation coefficient, R (P-value);  $s_w$ , CV% and ICC given with (95% CI).

Table 3. Typical sample size required to detect a small, medium and large effect size.

		Sample size (N)		
		0.2	0.5	0.8
During fatigue exercise	Endurance (s)	220	35	14
	$\Delta PC_{r_{ex}}$ (%)	424	68	26
	$\tau PC_{r_{ex}}$ (s)	1403	225	88
	TOI (%)	389	62	24
	$\Delta TOI$ (%)	494	79	31
	$\Delta HHb_{ex}$ (AU)	312	50	20
After 24 s exercise	$\tau PC_{r_{rec}}$ (s)	1082	173	68
	$\Delta HHb_{rec}$ (AU)	65	10	4
	$\tau HHb_{rec}$ (s)	838	134	52

The sample size is the total sample (N) required for comparing the change in a variable in an intervention group (N/2) with that in a control group (N/2).

Figure 1. Experimental set-up showing (a) the participant positioned prone on the scanner bed, (b) the location of the NIRS probe, (c) the location of the  $^{31}\text{P}$ -MRS coil, and (d) the participant holding an unsupported position.

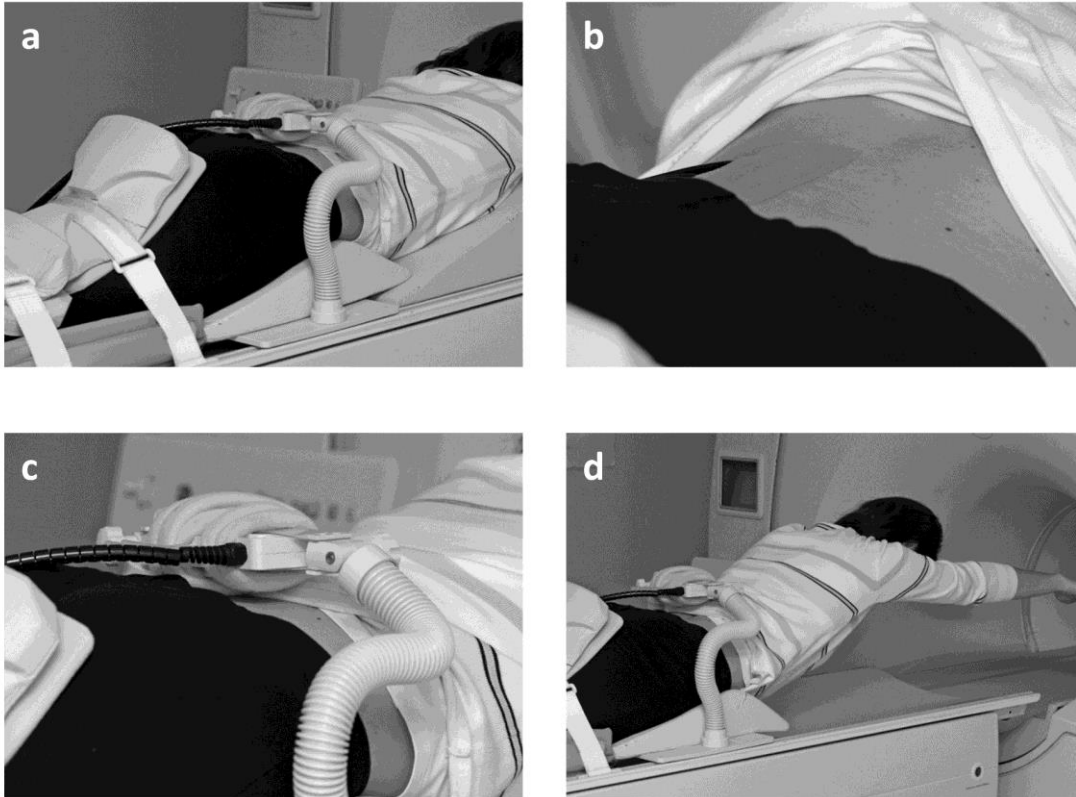


Figure 2. Scatter plot showing single-measures intra-class correlation coefficients against coefficients of variation for the eight  $^{31}\text{P}$ -MRS and NIRS measures.

