Reliability of a measurement method for the cross-sectional area of the longus colli using real-time ultrasound imaging

Cliona O'Riordan¹, Pepijn Van De Ven², John Nelson², Karen McCreesh¹ and Amanda Clifford¹

Ultrasound
2016, Vol. 24(3) 154–162
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1742271X16659099
ult.sagepub.com

Abstract

Objective: Real-time ultrasound imaging is an established objective outcome measurement with proven reliability. However, it is still largely biased by user-ability. Published research in the area of real-time ultrasound imaging reliability in the cervical region and in particular the deep cervical flexors is quite sparse. The purpose of this investigation was to examine if a novice ultrasound user could agree favourably with an experienced ultrasound sonographer in measuring the cross-sectional area of the longus colli.

Methods: Ultrasound images were captured from 22 healthy subjects on two different occasions, one week apart, by a novice ultrasound user. They were acquired using a GE Healthcare LOGIQe ultrasound machine, at a depth of 3–4 cm with the transducer frequency set to 8 MHz, in line with previous research guidelines. Cross-sectional area was then measured on-screen by both a novice and experienced ultrasound user to determine inter-rater reliability. Intra-rater reliability was also analysed using cross-sectional area figures from days one and two.

Results: Intra-rater reliability for real-time ultrasound imaging for the cross-sectional area of the longus colli was “excellent” (intra-class correlation 0.90, 95% CI 0.82–0.95). Inter-rater reliability was “moderate” but in keeping with previous published research (intra-class correlation 0.61, 95% CI 0.37–0.77).

Conclusion: Difficulties in identifying the borders of the longus colli muscle due to its deep anatomical location and surrounding structures make it difficult for assessors to agree favourably on cross-sectional area measurements, leading to “moderate” levels of inter-rater reliability and poor agreement. Intra-rater reliability is excellent, and in this instance indicates that a novice user can be just as reliable as a more experienced ultrasound user.

Keywords

Real-time, ultrasound, longus colli, reliability

Date received: 16 February 2016; accepted: 16 June 2016

Introduction

The deep cervical flexors, namely longus colli (Lco) and capitis, function to stabilise the cervical spine during dynamic activity, due to their direct cervical vertebral attachment.¹ Similar to the dysfunction identified in the lumbar paraspinal muscles in people with chronic lower back pain,² altered function and weakness have been identified in the deep cervical flexor muscles in those presenting with chronic neck pain (CNP).^{3–5} Up to

¹Department of Clinical Therapies, University of Limerick, Limerick, Ireland

²Department of Science and Engineering, University of Limerick, Limerick, Ireland

Corresponding author:

Cliona O'Riordan, Department of Clinical Therapies, Health Sciences Building, 2nd Floor, University of Limerick, Limerick, Ireland.
Email: Cliona.ORiordan@ul.ie

70% of the population experience neck pain at some point in their lives,⁶ with approximately 20% of these going on to experience persistent pain with frequent or daily recurring symptoms.^{3,7} Thus, it is important for clinicians to objectively assess these muscles to ascertain if weaknesses exist and target them in an individually tailored treatment programme.

Due to the deep location of the longus colli (Lco) and capitis, traditional methods of assessment, such as palpation, are difficult to perform accurately. Manual muscle testing of the deep cervical muscles in isolation is also difficult to execute without recruiting the superficial cervical muscle layer.⁸ Techniques to specifically isolate the deep cervical flexors using fine electromyography wires, as developed by Falla,⁹ whilst effective and highly accurate, are invasive, uncomfortable and clinically inapplicable. Objective methods such as real-time ultrasound (RT-US) and magnetic resonance imaging (MRI) have become popular, both clinically and for research purposes, to assess and measure muscle characteristics including cross-sectional area (CSA), muscle fibre pennation, fascicle length and CSA.¹⁰ CSA is a useful measure, as previous studies have shown a strong correlation between muscle CSA and muscle strength.^{11,12}

MRI and RT-US have proven reliability for use on a variety of muscle groups¹³ to provide an objective measure of muscle atrophy or hypertrophy. Muscle impairment may persist after pain symptoms have ceased, and specific muscle re-training may be required. MRI and RT-US can be used to assess muscle activity, atrophy, hypertrophy and activation reliably if required.^{14,15} MRI provides multi-planar images; however, its disadvantages include its high cost and limited accessibility, as well as limited real-time capacity and obvious reduced patient group contraindications (e.g. claustrophobia, metallic implants, pacemakers and pregnancy). In comparison, real-time ultrasound is cost-effective, accessible, clinically applicable, and portable.¹⁶ In addition, RT-US can help guide rehabilitation through its ability to provide instantaneous feedback in real-time for clinicians and can be used to capture dynamic images, thus allowing observation of movement patterns, muscle contractile abilities, activation deficits, muscle recruitment patterns, and morphological changes.^{13,17,18}

Unlike MRI, the quality of the image obtained from a real-time ultrasound machine is highly dependent on the experience of the user and can be viewed as a major disadvantage.¹³ RT-US has been used extensively with “excellent” intra-rater reliability, as well as “good” inter-rater reliability and validity in larger groups of muscles such as the quadriceps and hamstrings complexes and in the lumbo-pelvic region;¹⁹ however, its use at cervical spine level and in associated musculature is sparse.^{8,14} Previous research examining the reliability

of ultrasound imaging for measuring the CSA of the Lco has varied in terms of methodology and consequent results. Intra-rater reliability figures varied, with intra-class correlation (ICC) values ranging from 0.71 to 0.90, whilst inter-rater reliability values ranged from 0.68 to 0.82.^{8,18,20} Inter-rater reliability of Lco CSA measurements for inexperienced RT-US users ranged from 0.68 (95% CI 0.44–0.87)¹⁸ to 0.81 (95% CI 0.64–0.90).⁸ In a study by McGaugh and Ellison,¹⁸ the physiotherapists only received three hours of RT-US tuition from an experienced user, resulting in a lower inter-rater reliability. Whilst the examiners used in the study by Cagnie et al.⁸ were more experienced, this was reflected in the subsequently higher inter-rater reliability co-efficient, thus identifying the overwhelming influence of experience on RT-US use.

Given the limited published research in this area, this current study was undertaken in a healthy population to determine if a novice ultrasound user with limited experience could agree favourably with an experienced ultrasound practitioner using an on-screen measurement method. Thus, the objectives of this current study were (1) to determine the intra-rater reliability of an ultrasound imaging protocol for measuring the CSA of the longus colli in a healthy population and (2) to determine the inter-rater reliability between a novice ultrasound user and an experienced ultrasound practitioner using an RT-US imaging measuring protocol for the longus colli.

Materials and methods

Recruitment and subjects

Volunteers were recruited from a sample of convenience within a university population for this ethically approved reliability study (Ethics number 2012_11_01_EHS). Subject demographics including age, weight and height are shown Table 1. Subjects were included if they reported an absence of neck pain or cervicogenic headaches in the six weeks prior to testing and at the time of testing.⁸ Exclusion criteria included a history of neck or back trauma, neurological and/or inflammatory disorders, dizziness, vestibular symptoms and allergies to ultrasound gel.

Study design and procedure

Ultrasound images were captured using a GE Healthcare (Fairfield, Connecticut, USA) LOGIQ e ultrasound machine with a 12 MHz linear array probe set to a frequency of 8 MHz and at an average depth of 3–4 cm, in line with previous research guidelines.^{8,18} These relaxed state ultrasound images were acquired by a physiotherapist, a novice RT-US user, on two

Table 1. Subject demographics for the 22 participants

	Women (n = 15)	Men (n = 7)	Overall (n = 22)
Age (years) (mean $\pm$ SD)	22.27 $\pm$ 4.5	35 $\pm$ 10.86	26.23 $\pm$ 9.14
Weight (kg) (mean $\pm$ SD)	64.33 $\pm$ 6.3	86 $\pm$ 24.15	70.86 $\pm$ 14.46
Height (m) (mean $\pm$ SD)	167.13 $\pm$ 8.28	178.82 $\pm$ 13.7	171.28 $\pm$ 10.71
CSA Lco (cm ²) (mean and SD)	0.65 $\pm$ 0.17	0.65 $\pm$ 0.17	0.67 $\pm$ 0.16

SD: standard deviation.

Figure 1. Imaging protocol and probe placement for longus colli.

test occasions, one week apart. The novice user had received tuition from Assessor 2 (an experienced and qualified RT-US sonographer) and had conducted independent practice using RT-US for a minimum of 10 hours. On-screen CSA measurements of the Lco muscle were then made independently by both assessors at different times using the measure mode setting on the GE Healthcare LOGIQ e ultrasound machine. One CSA measurement of each Lco image was made by each assessor using the “roller-ball” mouse on the machine to create a continuous trace of the muscle borders. Both assessors were blinded to the other assessor’s on-screen measurements until data were accumulated for statistical analysis.

Ultrasound imaging technique

Subjects lay supine with a rolled towel positioned posteriorly to flatten the cervical lordosis.²⁰ The thyroid cartilage was identified by placing the probe perpendicular to the longitudinal axis of the anterior aspect

Figure 2. RT-US image of the deep cervical anatomy, including landmarks for anatomical referencing. SCM: sternocleidomastoid; Thy: thyroid; CA: carotid artery; Lco: longus colli; IJV: internal jugular vein; PVF: prevertebral fascia.

of the neck. The probe was then moved distally approximately 2 cm; this corresponds to the level of the fifth and sixth cervical vertebrae.^{8,18} This level is chosen as it permits clear visualisation of the longus colli, where the muscle attaches laterally to the anterior tubercle of C6, and at this level it is not overlapped by longus capitis.⁸ The probe was moved 1 cm laterally left and right of the lower laryngeal prominence of the thyroid cartilage.²⁰ Doppler mode was used to verify the correct location by identifying the carotid artery and, inferio-medially, the Lco (see Figure 1).¹⁸

Anatomical landmarks used to identify the Lco muscle borders included: the carotid and internal jugular vein (antero-laterally), the thyroid cartilage (antero-medially) with the recto-pharyngeal space superiorly as per previous research protocols (see Figure 2).^{8,14}

Statistical analysis

CSA measurements for left and right Lco muscles were collected from both assessors and entered

into SPSS Version 20.0 for statistical analysis. Both left and right CSA images of the Lco were taken on two occasions from 22 subjects; thus, 44 images were selected for each analysis. Reliability analyses, including ICC equations (1,1) for intra-rater reliability and (2,1) for inter-rater reliability were conducted on CSA measurements from days 1 and 2 and Day 2 only, respectively. For each pair of CSA measurements from Day 1 and Day 2, the difference and average were calculated to construct the Bland and Altman graphical representation for intra-rater reliability.²¹ Similar methods were used for Day 2 images to calculate the average and difference in CSA measurements between raters for inter-rater reliability. In accordance with Shrout and Fleiss,²² ICC correlation coefficients were deemed “excellent” (>0.75), “moderate” (>0.50 <0.74) or “poor” (<0.49).

Data analysis

The standard error of a measurement (SEM) quantifies the extent to which a test provides accurate scores. A low SEM indicates high levels of score accuracy and vice versa.

$$\text{SEM} = \text{standard deviation from the 1st test} \\ \times \left(\sqrt{1 - \text{ICC}} \right)$$

Minimal detectable change (MDC) indicates the magnitude of change below which there is more than 95% chance that no real change has occurred.

$$\text{MDC} = 1.96 \times \text{SEM} \times \sqrt{2}$$

Results

Intra-rater reliability

The ICC coefficient for intra-rater reliability was “excellent”,²² with a value of 0.90 (see Table 2). The Bland and Altman²¹ plot shown in Figure 3 demonstrates small differences between measures and narrow

limits of agreement (−0.17 to 0.15), which indicates favourable levels of agreement between measurements. The mean difference is small (−0.01 cm²) indicating little bias between measurements.

Inter-rater reliability

The ICC coefficient for inter-rater reliability (0.61, see Table 2) was moderate according to the Shrout and Fleiss definition.²² The Bland Altman plot in Figure 4 shows that the mean difference between measures (0.26 cm²) was larger than that for intra-rater reliability, indicating a bias between observers. Assessor 2 systematically reported larger measurements than assessor 1. This bias is also reflected in the limits of agreement, which were −0.15 to 0.67. The limits of agreement are much wider than for the intra-rater study, showing poorer agreement between the two assessors. Only one case lay outside the 95% limits of agreement as seen in Figure 4.²¹

Standard error of the measurement and MDC for intra and inter-rater reliability

Standard error of measurement figures and MDC figures are given in Table 3. Based on figures for intra-rater reliability, an MDC of 0.30 cm² of the Lco muscle would be the smallest change in size that could be detected beyond random error. For inter-rater reliability, this is even larger at 0.75 cm²; these figures are dependent on the sample size distribution.

Discussion

Intra-rater reliability

The intra-rater reliability for real-time ultrasound measurement of the CSA of the Lco in a healthy population was excellent (ICC 0.90, 95% CI 0.83–0.95). This is comparable to available published literature for ultrasound of the Lco and is higher than intra-rater reliability figures by McGaugh and Ellison¹⁸ and

Table 2. Inter and intra-rater reliability statistics for both ICC coefficient and Bland–Altman methods

	Interclass correlation coefficient		Bland–Altman method (cm ²)				
	ICC	95% CI	d	SE of d	95% CI for d	SD (difference)	Limits of agreement
Inter-rater reliability	0.61	0.37–0.77	0.26	0.03	0.19–0.33	0.21	−0.15–0.67
Intra-rater reliability	0.90	0.83–0.95	0.01	0.01	−0.30–0.02	0.08	−0.17–0.15

ICC: intra-class correlation co-efficient; 95% CI: 95% confidence intervals; d: mean difference; SE of d: standard error of the mean difference; 95% CI for d: 95% confidence interval of the mean difference; SD: standard deviation.

Figure 3. Bland and Altman²¹ plot showing mean difference and limits of agreement for intra-rater reliability of ultrasound measurement of the longus colli cross-sectional area in a healthy population.

-----=upper and lower limits of agreement.

_____ = mean difference.

Javinsir et al.²⁰ who reported ICC values ranging from 0.67 and 0.87. Possible causes of this may be due to the combination of anatomical referencing and the use of Doppler ultrasound in identifying the correct location in this study. Published research for the reliability of ultrasound imaging for measuring the CSA of the LCo in both healthy and neck pain populations has varied in terms of methodology and consequent results. Intra-rater reliability figures ranged from 0.71 to 0.90.^{8,18,20} Levels of experience of assessors ranged from novices to experienced; however, all had received some level of training in ultrasound imaging of the longus colli. The methods of measurement, frequency used and field depth, as well as the equipment used, including probe type, may contribute to the differences observed in results between this current study and previous research.

The limits of agreement (−0.17–0.15) suggest a narrow margin of variability in measurements taken between Day 1 and Day 2 (Table 2). The low value of mean difference (−0.01) indicates excellent agreement and little variability in measurements between test occasions.²³ Thus, this confirms that the current protocol

utilised in this study is reliable in a healthy population over two testing occasions for a novice ultrasound user. Hence, using a combination of anatomical reference points and the Doppler mode, if available on the ultrasound machine, is an effective means of increasing standardised imaging and reliability in measuring the LCo on different test occasions by the same assessor.

Inter-rater reliability

ICC values for inter-rater reliability in this current study were “moderate” (ICC 0.61; 95% CI 0.38–0.77), but in line with other research.¹⁸ Cagnie et al.⁸ previously demonstrated moderate inter-examiner reliability (ICC 0.68; 95% CI 0.48–0.81) for measurements of the LCo in healthy subjects. Javanshir et al.¹⁴ have demonstrated within-day and between-day reliability of between 0.71 and 0.81 for subjects with mechanical neck pain, showing that real time ultrasound is a reliable tool for patients with neck pain as well as healthy populations. This finding in the current study may be attributable to a combination of factors. The wide

Figure 4. Bland and Altman²¹ plot showing mean difference and limits of agreement for inter-rater reliability of ultrasound measurement of the longus colli cross-sectional area in a healthy population.

-- = upper and lower limits of agreement.

_____ = mean difference.

Table 3. Standard error of measurement and minimal detectable change for intra and inter-rater reliability

	SEM [cm ²]	MDC [cm ²]
Intra-rater reliability	0.11	0.30
Inter-rater reliability	0.27	0.75

SEM: standard error of the measurement; MDC: minimal detectable change.

limits of agreement (−0.15–0.67) between assessors show large variability for inter-rater reliability of ultrasound for this particular muscle. The intervals are wide reflecting the small sample size (<50) and the great variation in differences.²¹ The mean difference was also quite large (0.26 cm²) and not as close to zero as the authors would have desired (Table 2). This would indicate a systematic bias and that the experienced assessor (Assessor 2) produced larger measurements compared to Assessor 1. Figures 5 and

6 demonstrate best agreement and largest discrepancy for CSA measurements of the Lco between both assessors. As can be seen, measurement differences ranged from 0.01 cm² to 0.67 cm². It is important to note that variability in measurement may have been influenced by the novice assessor (Assessor 1) being able to observe the image and muscle borders in real-time. However, due to the study methodology, this was not possible for Assessor 2. As the inter-rater reliability can only be classed as “moderate”, it is important to consider the reasons for differences between this current study and previous research in the area. McGaugh and Ellison¹⁸ found smaller mean differences for inter-rater reliability (0.04 cm²) and narrower limits of agreement (±0.23 cm²). Both assessors in that study had similar levels of training in ultrasound imaging and both were present at the time of image collection. A similar ultrasound machine was used in this procedure; however, different frequency settings were utilised, which may have contributed to different results found between studies. As can be seen in Table 3,

Figure 5. Ultrasound images demonstrating on-screen CSA measurements by (a) Rater 1 (expert) and (b) Rater 2 (novice), where agreement was most favourable. For these particular images, there was a difference of 0.02 cm^2 between measurements.

a large standard error of measurement for inter-rater reliability was calculated and a corresponding large MDC of 0.75 cm^2 . This is in keeping with the wide limits of agreement obtained for between-rater measurements.

In the case of inter-rater reliability, agreement on identification of the muscle borders may account for the “moderate” reliability found in this current study. In particular, the lateral border can cause great difficulty due to the indistinct outline of the muscle, thus making verification variable. The posterior border can be difficult to delineate due to the acoustic shadow of the trachea.¹⁴

CSA measurements

When looking at exact figures of CSA of the Lco in this study, the average CSA was $0.67 \pm 0.16 \text{ cm}^2$. This is substantially smaller than the average CSA values

Figure 6. Ultrasound images demonstrating on-screen CSA measurements by (a) Rater 1 (expert) and (b) Rater 2 (novice), where agreement was not favourable. For this particular image, there was a difference of 0.42 cm^2 between measurements. Distinguishing the posterior border of the muscle in this image, as mentioned in the Discussion is an area for discrepancy between raters.

(e.g. controls $0.85 \pm 0.11 \text{ cm}^2$ and neck pain patients $0.66 \pm 0.11 \text{ cm}^2$) measured in previous literature.¹⁴ Similarly, Cagnie et al.⁸ reported an average CSA of $1.22 \pm 0.27 \text{ cm}^2$ in healthy controls. The differences observed, as mentioned above, may be due to the different ultrasound machines utilised in each study, the method by which the CSA was measured either on-screen or using an alternative software program. The majority of this group were also females ($n = 15$) and as Javinshir et al.²⁰ demonstrated, there is an effect of sex on CSA of muscle size. Findings from an ANCOVA showed a significant difference in the CSA of the Lco between men and women ($p = 0.018$) in healthy people only. This may also account for the variances seen in CSA figures between this study and previously published research.

Limitations

Conflicting interpretations of muscle borders and anatomical reference points by different assessors may have led to the over- or under-estimation of the muscle borders and the subsequent CSA measurements for both intra and inter-rater reliability.¹³ As the aim of this current study was to determine the reliability of an onscreen measurement protocol between two assessors, it is important to note that this measurement involved the use of a roller mouse. Varying abilities in the use of this type of control may have influenced the disparities in measurements calculated for both intra and inter-rater reliability statistics. This should be carefully considered in interpreting the results of this current study.¹⁸

As with all reliability studies, it should be noted that regardless of the reliability tests conducted and subsequent results, the comparison of reliability results between studies is limited unless the size and attributes of the sample populations tested in each case are extremely similar.²³

Possible clinical applications

This protocol was used in a healthy population; future research is required to investigate if similar or improved reliability can be achieved in a CNP clinical population. Evidence suggests that the deep neck flexors of those with neck pain are impaired and appear more marbled, i.e. containing more fatty infiltrate than healthy controls, making border identification more difficult.¹⁸ Thus, it is important to have excellent reliability in a healthy population, to aim to be equally reliable in a neck pain population. Additional statistical analyses were conducted to find the SEM of images in this current study to provide evidence of precision of measurements taken. For intra-rater reliability, this was 0.11 cm²; this low value indicates how accurate the measurements were. MDC was also calculated; this gives an indication of the minimal change required for that change to be attributable to an intervention, in this case 0.30 cm². This figure is quite high, given the small CSA of the muscle.

As can be seen from this current study, intra-rater reliability of Lco measurement for a novice real-time ultrasound user is "excellent". However, the reliability for on-screen CSA between two assessors is only moderate. RT-US was more reliable when one individual is conducting re-assessments, thus clinically, it is more advantageous if one clinician can perform all assessments. This may not be always possible, and whilst it is considered an objective measure, it is still largely dependent on the user.¹³

Conclusion

Assessments of the deep cervical flexors, which are impaired in neck pain, are difficult to execute reliably in a clinical setting, but are a requirement in a chronic neck pain population. Real-time ultrasound is a feasible and evidently reliable method to conduct such an objective assessment of muscle function for an individual assessor but less so between two assessors. Further research to determine if this current study protocol is as reliable in a chronic neck pain population is required, as well as determining if a significant difference exists between the CSA of the Lco in healthy controls and those with neck pain.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical approval

Local ethical approval was obtained for this reliability study (Ethics number 2012_11_01_EHS).

Guarantor

COR.

Contributorship

COR researched the literature and conceived the study, and wrote the final version of the manuscript. AC, KMC assisted COR in the methodology design. KMC acted as Assessor 2 for the reliability study. AC, KMC, PVdV, JN provided revisions for draft manuscripts. All authors reviewed and approved the final version of the manuscript.

References

1. Falla D, Jull G, Hodges P, et al. An endurance-strength training regime is effective in reducing myoelectric manifestations of cervical flexor muscle fatigue in females with chronic neck pain. *Clin Neurophysiol* 2006; 117: 828–837.
2. Moseley GL and Flor H. Targeting cortical representations in the treatment of chronic pain: a review. *Neurorehabil Neural Repair* 2012; 26: 646–652.
3. Falla D. Unravelling the complexity of muscle impairment in chronic neck pain. *Man Ther* 2004; 9: 125–133.
4. Kristjansson E. Reliability of ultrasonography for the cervical multifidus muscle in asymptomatic and symptomatic subjects. *Man Ther* 2004; 9: 83–88.
5. Jull G, Falla D, Vicenzino B, et al. The effect of therapeutic exercise on activation of the deep cervical flexor

- muscles in people with chronic neck pain. *Man Ther* 2009; 14: 696–701.
6. Viljanen M, Uitti J, Palmroos P, et al. Effectiveness of dynamic muscle training, relaxation training, or ordinary activity for chronic neck pain: randomised controlled trial. *BMJ* 2003; 327: 475–477.
 7. Ehrlich G. Low Back Pain. *Bull World Health Organ* 2003; 81: 671–676.
 8. Cagnie B, Derese E, Vandamme L, et al. Validity and reliability of ultrasonography for the longus colli in asymptomatic subjects. *Man Ther* 2009; 14: 421–426.
 9. Falla D. Patients with chronic neck pain demonstrate altered patterns of muscle activation during performance of a functional upper limb task. *Spine* 2004; 29: 436–440.
 10. Bemben M. Use of diagnostic ultrasound for assessing muscle size. *J Strength Cond Res* 2002; 16: 103–108.
 11. Maughan RJ, Watson JS and Weir J. Strength and cross-sectional area of human skeletal muscle. *J Physiol* 1983; 338: 37–49.
 12. Maughan RJ, Watson JS and Weir J. Muscle strength and cross-section area in man: a comparison of strength-trained and untrained subjects. *Br J Sports Med* 1984; 18: 149–157.
 13. Whittaker J, Tayhen D, Elliott J, et al. Rehabilitative ultrasound imaging: understanding the technology and its application. *J Orthop Sports Phys Ther* 2007; 37: 434–449.
 14. Javanshir K, Amiri M, Mohseni-Bandpei MA, et al. Ultrasonography of the cervical muscles: a critical review of the literature. *J Manipulative Physiol Ther* 2010; 33: 630–637.
 15. Tsao H and Hodges P. Immediate changes in feedforward postural adjustments following voluntary motor training. *Exp Brain Res* 2007; 18: 537–548.
 16. Campbell S, Adler R and Sofka C. Ultrasound of muscle abnormalities. *Ultrasound Q* 2005; 21: 87–94.
 17. Henry S and Westervelt K. The use of real-time ultrasound feedback in teaching abdominal hollowing exercises to healthy subjects. *J Orthop Sports Phys Ther* 2005; 35: 338–345.
 18. McGaugh J and Ellison J. Intrasection and interrater reliability of rehabilitative ultrasound imaging measures of the deep neck flexors: a pilot study. *Physiother Theory Pract* 2011; 27: 572–577.
 19. Van K, Hides J and Richardson C. The use of real-time ultrasound imaging for biofeedback of lumbar multifidus muscle contraction in healthy subjects. *J Orthop Sports Phys Ther* 2006; 36: 920–925.
 20. Javanshir K, Mohseni-Bandpei M and Amiri M. Ultrasound assessment of bilateral longus colli muscles in subjects with chronic bilateral neck pain. *Am J Phys Med Rehabil* 2011; 90: 293–301.
 21. Bland J and Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 327: 307–310.
 22. Shrout P and Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420–428.
 23. Rankin G, Stokes M and Newham DJ. Size and shape of the posterior neck muscles measured by ultrasound imaging: normal values in males and females of different ages. *Man Ther* 2005; 10: 108–115.