Ultrasound Obstet Gynecol 2014; **44**: 481–485
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.13409

Measuring echogenicity and area of the puborectalis muscle: method and reliability

A. T. M. GROB*†#, A. A. C. VEEN†#, K. J. SCHWEITZER*, M. I. J. WITHAGEN*, G. A. VAN VEELEN* and C. H. VAN DER VAART*

**Department of Reproductive Medicine and Gynecology, University Medical Center, Utrecht, The Netherlands; †MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands*

KEYWORDS: 3D/4D ultrasound; delineation; echogenicity; pelvic floor; puborectalis muscle; reliability

ABSTRACT

Objectives *To develop a semi-automated method to assess puborectalis muscle echogenicity on three-dimensional/four-dimensional (3D/4D) volume transperineal ultrasound images using 4D View and Matlab® software and evaluate its intra- and interobserver reliability.*

Method *The data of 23 women in their first trimester were included. 3D/4D volume datasets were obtained at rest. Two inexperienced observers were trained by an experienced observer to construct tomographic ultrasound images (TUI) from the original data and to delineate all structures. Puborectalis muscle area (PMA) and the mean echogenicity of the puborectalis muscle (MEP) were calculated offline. Intra- and interobserver reliability were determined by intraclass correlation coefficients (ICC) and their 95% CIs.*

Results *The development of a semi-automated method to calculate puborectalis area and echogenicity is described in detail. PMA and MEP measurements in pregnant women demonstrated almost perfect intraobserver reliability for both inexperienced observers, with ICC values ranging from 0.88 to 0.99. The interobserver reliability showed ICCs of 0.63 for PMA and almost perfect ICC values, of 0.96–0.98, for echogenicity. The majority of intraobserver mismatch between two delineations of PMA occurred near the borders.*

Conclusions *Matlab software can be used to provide reliable measurements of the area and echogenicity of the puborectalis muscle. As the latter can be used to assess structural changes in the puborectalis muscle, it appears a promising new tool for studying pelvic floor structural anatomy. Copyright © 2014 ISUOG. Published by John Wiley & Sons Ltd.*

INTRODUCTION

In recent years, three-dimensional (3D) and four-dimensional (4D) volume transperineal ultrasound imaging have become increasingly popular for studying pelvic floor anatomy, allowing visualization of the pelvic floor in the axial plane[1]. The axial plane is used to measure dimensions of the pelvic floor anatomy, such as the area of the puborectalis muscle, area of levator hiatus and minimal hiatal distances[2]. Although the measurement of dimensions of the pelvic floor is well developed, identifying structural changes in the puborectalis muscle, apart from levator avulsions, is still in its infancy[2–5]. One option to assess muscular structure in a more quantitative way is to measure its echogenicity[6,7]. Echogenicity measurements are already being used in the diagnosis and evaluation of neuromuscular disorders, as well as in orthopedics[8,9]. This study was designed to develop and test the reliability of a semi-automated method to measure mean echogenicity of the puborectalis muscle (MEP).

METHOD

Development of our method to measure echogenicity was carried out as part of a subanalysis of a large study in our University Hospital. Over a period of 2 years, 280 nulliparous pregnant women were seen for 3D/4D transperineal ultrasound assessment of their pelvic floor anatomy during and after pregnancy. For our research question we used the 4D ultrasound data subsets of 23 randomly chosen women with a singleton pregnancy at approximately 12 weeks' gestation. Women were excluded if they had a medical history of urinary and/or fecal incontinence, previous prolapse or anti-incontinence surgery, connective tissue disease or

Correspondence to: A. T. M. Grob, Department of Reproductive Medicine and Gynecology, University Medical Center, PO Box 85500, 3508GA Utrecht, The Netherlands (e-mail: a.t.m.grob-3@umcutrecht.nl)

#A. T. M. Grob and A. A. C. Veen contributed equally to this work.

Accepted: 2 May 2014

neurological disorders. The Institutional Human Research Ethics Committee approved the study and all women gave informed consent.

Sonographic assessment consisted of 4D transperineal ultrasound imaging using a GE Voluson 730 Expert system (GE Medical Systems Zipf, Austria) with a RAB 4–8-MHz curved array volume transducer. The angles of the acquired volume were set 85° longitudinal and 70° transverse to the probe, and the depth of the volume varied per measurement. A temporal resolution of 3 Hz was used to acquire the data, and settings that could influence the intensity values were kept constant for each measurement. These settings were: gain, 15; power, 100; Harmonics, mid; contrast, 8; gray map, 4; persistence, 8; and enhance, 3. All pelvic floor ultrasound examinations were performed with the participants supine and with an empty bladder[2]. The ultrasound probe was placed on the perineum in the sagittal plane and measurements made with musculature at rest were used for our analysis. The datasets were stored on a hard disk for analysis offline.

Offline analysis of the data was performed using 4D View 7.0 (GE Medical Systems) and Matlab® R2010a (MathWorks, Natick, MA, USA) by two inexperienced observers (A.G. and A.V.) and one experienced observer (K.S.). The inexperienced observers were trained in two sessions, covering 20 cases, by the experienced observer. Image analysis was performed by first determining and fixating the point of time of total muscle relaxation (4D data turned to 3D data). The plane of minimal hiatal dimensions was selected as previously described[2,10,11]. This plane was used to obtain tomographic ultrasound images (TUI) in the axial direction. The first slice in which the symphysis seemed closed was used for analysis[5,10]. This two-dimensional (2D) ultrasound image contained 1304×662 pixels and was exported as a .bmp file to Matlab R2010a (Image Processing Toolbox 7.0).

Figure 1 shows the semi-automated method used to select the puborectalis muscle and levator hiatus in three steps. The Matlab function 'imfreehand' was used for delineation. First, an outer border around the area of interest, consisting of the puborectalis muscle and levator hiatus, was drawn (Figure 1, Step 1a). All data outside this area of interest were eliminated and turned black (Figure 1, Step 1b). The second step was to draw a line to select the levator hiatus. This line followed the inner border of the puborectalis muscle, pubic symphysis and inferior pubic ramus (Figure 1, Steps 2a and 3a). The third step was to select the puborectalis muscle by drawing two lines at the attachment of this muscle to the symphysis (Figure 1, Step 3b). The resulting image represents the area of the puborectalis muscle that is automatically calculated in cm^2. Differences in measurements may occur if markers are positioned in areas with less well-defined demarcation. In order to analyze this mismatch area we obtained an overlay of two delineation attempts on the same puborectalis muscle image. When a pixel was included in a delineation on both attempts, the pixel turned black, whereas if the pixel was delineated in only one of two attempts, the pixel retained its original (gray) color. The qualitative analysis is based on identifying the largest areas of mismatch between two delineation attempts. Quantitative analysis was performed by dividing the area of the mismatch (in pixels) by the total puborectalis muscle area (PMA) (in pixels).

Determination of echogenicity was based on the gray-scale image, in which the value for each pixel could range from 0 (black) to 255 (white). Normal muscle cells are rather echolucent and appear dark on the image. The connective and fatty tissues of the muscle have a higher echogenicity and appear brighter[12]. The mean of all the pixel echogenicity values of the puborectalis muscle was calculated automatically, and is referred to subsequently as MEP.

The reliability of measuring PMA and MEP was tested in both intra- and interobserver series, with three independent examiners analyzing the patient's randomly ordered datasets. There was a 4-day time window between the repeat measurements and analysis of the offline dataset. At the time of the second delineation, the observers were blinded to the outcome of the first delineation and the datasets were again ordered randomly.

Statistical analysis

Statistical analysis was performed using SPSS v. 20 (SPSS Inc., Chicago, IL, USA) and Excel 2010 (Microsoft Office, Microsoft Corp, Redmond, WA, USA). Means, SD and intraclass correlation coefficients (ICCs) with their 95% CIs were used to compare the datasets and validate the delineation. The ICC results were classified according to the subgroups defined by Landis and Koch, in which ICC values below 0.00 were considered poor, 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1.00 almost perfect[13]. To evaluate the mean difference and limits of agreement (LOA) between observers, the Bland–Altman method was used[14]. A 95% CI for the bias was used to test for significance, to verify that the bias did not differ from zero.

RESULTS

The mean age of the women was 30.9 (SD 3.8) years and their mean body mass index (BMI) was 23.3 (SD 4.3). Means with SD, ICC values with 95% CIs and mean differences with limits of agreement for PMA and MEP are shown in Table 1. Additionally, in Figure 2 the Bland–Altman LOAs for PMA measurements are shown.

PMA and echogenicity

The PMA had an almost perfect intraobserver (ICC: 0.88–0.94) and a moderate to almost perfect interobserver (ICC: 0.63–0.87) reliability. The MEP measurement showed almost perfect intra- and interobserver ICCs (range, 0.96–0.99).

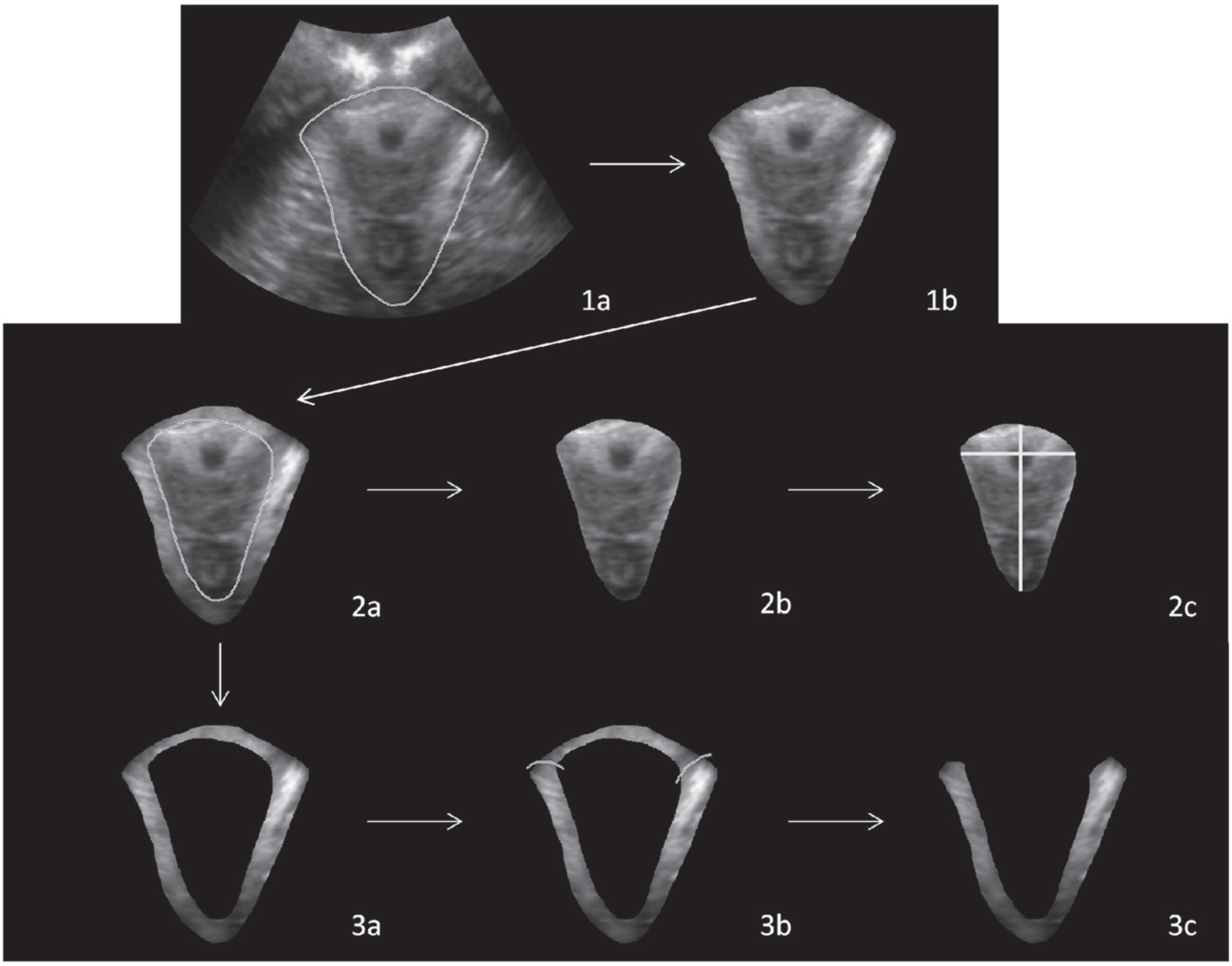

Figure 1 Semi-automated method using three steps (Steps 1–3) to select the puborectalis muscle and levator hiatus for assessment of area and echogenicity. Steps 1a and 1b: the area of interest is delineated. Steps 2a and 2b: the levator hiatus is delineated. Step 2c: minimal hiatal dimensions are visualized. Steps 3a–3c: the puborectalis muscle is delineated.

Table 1 Intra- and interobserver reliability of delineation of pelvic floor musculature at rest

	Intraobserver reliability			*Interobserver reliability*			
Parameter	*Observer*	*Mean (SD)*	*ICC (95% CI)*	*Observer*	*ICC (95% CI)*	*Mean difference (SD)*	*LOA (lower–upper)*
PMA (cm^2)	A.G.	8.0 (1.6)	0.91 (0.80–0.96)	A.G./A.V.	0.69 (0.39–0.85)	−1.17 (1.56)	−4.29 to 1.96
	A.V.	9.2 (2.3)	0.88 (0.72–0.95)	A.G./K.S.	0.87 (0.71–0.94)	−0.08 (0.94)	−1.96 to 1.80
	K.S.	8.1 (2.0)	0.94 (0.85–0.97)	A.V./K.S.	0.63 (0.31–0.84)	1.08 (1.86)	−2.63 to 4.80
Mean echogenicity	A.G.	128 (19)	0.99 (0.96–0.99)	A.G./A.V.	0.96 (0.88–0.98)	2.53 (5.0)	−7.49 to 12.55
	A.V.	126 (19)	0.98 (0.96–0.99)	A.G./K.S.	0.97 (0.93–0.99)	1.76 (4.0)	−6.27 to 9.78
	K.S.	126 (19)	0.99 (0.97–0.99)	A.V./K.S.	0.98 (0.94–0.99)	−0.77 (4.2)	−9.12 to 12.55

Observers: A.G. and A.V., inexperienced; K.S., experienced; ICC, intraclass correlation coefficient; LOA, limits of agreement; PMA, puborectalis muscle area.

Area of mismatch

Figure 3 shows the differences in puborectalis muscle delineation caused by a mismatch in marker positioning. The left and center images show the two separate measurements and the right shows the mismatch. The mismatch tended to occur at the outer border of the puborectalis muscle. The total area of the mismatch ranged from 2 to 12%.

DISCUSSION

Our study shows that the area of the puborectal muscle and its echogenicity can be measured reliably using 3D/4D pelvic floor ultrasound combined with Matlab software. Inexperienced observers were able to perform these measurements adequately after 20 training sessions.

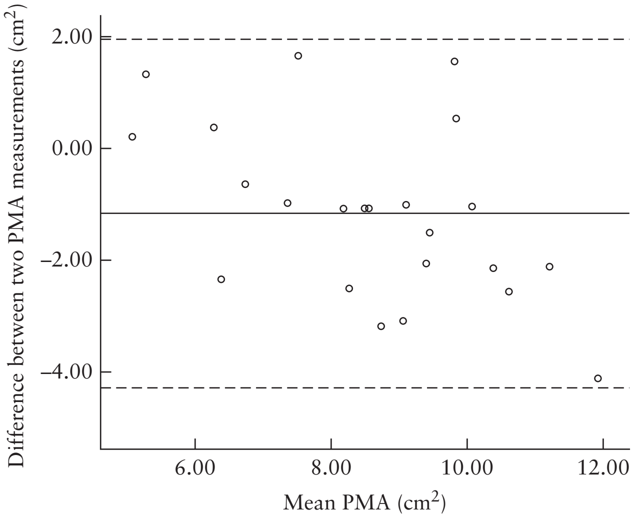

Figure 2 Bland–Altman analysis of puborectalis muscle area (PMA) measurements by Observer A.G., showing mean bias and limits of agreement.

To appreciate our findings, some possible limitations need to be discussed. First of all, we included healthy pregnant nulliparous women without previous delivery trauma to their pelvic floor. This provided us with high-quality images of intact pelvic floor musculature, which might have improved the accuracy of measurements. Another limitation is that two inexperienced observers (A.G. and A.V.) were trained by the same experienced observer (K.S.), which might have introduced instructor bias. However, training was given according to universally accepted pelvic floor ultrasound image-analysis guidelines (TUI reconstruction). Studying gray-scale images introduces another possible drawback. Different settings, within or between the ultrasound system(s), may produce different gray-scale images. This issue has been recognized and conversion equations can be applied to address this problem[15]. In our study we used deliberately the same specific settings in all measurements to ensure that the changes we observed in MEP were not affected by this potential bias. Finally, although we blinded the observers to the first set of measurements, the limited sample size introduces the risk of recall bias. We tried to limit this by using a 4-day time window between the time of image delineation and a separate random analysis of the two image sets.

One of the major strengths of our study is that we introduced a new parameter, MEP, into the research area of studying pelvic floor anatomy. We showed that measurements of MEP can be performed reliably.

The interobserver reliability of measuring PMA had the lowest ICC value. As shown in Figure 3, this is most probably caused by mismatching the outer border of the puborectalis muscle. Clear landmarks, such as dark edges indicating the border, were often not found, forcing the observers to delineate the structure more arbitrarily. This resulted in an average difference of 300–2000 pixels between measurements. Relative to the total number of pixels in the PMA, this accounts for 2–12% of the total area. However, the almost perfect MEP ICC values show that this mismatch did not affect the mean echogenicity measured.

Adding computer software to the delineation process of structures in the pelvic floor decreases the likelihood of human error and might also result in a reduction in the time taken for the procedure as all parameters can be calculated from one Matlab cycle instead of from multiple separate drawings produced by 4D View software. Further (quantitative) research should indicate how much time can be saved by using our method. Applying the software requires limited experience with pelvic floor image interpretation and delineation. However, currently we still need to use two separate software systems (ultrasound and Matlab).

The semi-automatic detection method offers the possibility to study new parameters, such as echogenicity. Three recent studies examining the link between echogenicity and clinical parameters were performed by Tsai *et al.*[8], Maurits *et al.*[16,17] and Pillen *et al.*[18]. Tsai *et al.* demonstrated that a decrease in the mean gray-level (echogenicity) may be used as a sonographic indicator of rotator cuff partial-thickness tear or tendinopathy

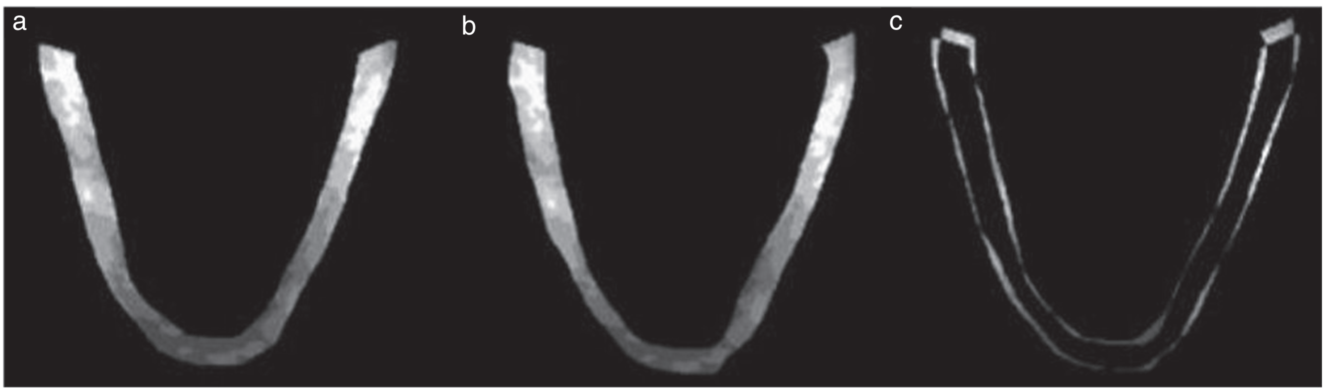

Figure 3 Qualitative analysis of the largest area of mismatch (c) between two delineation attempts (a and b). The mismatch occurs at the outer border of the puborectalis muscle and the total area of mismatch ranged from 2% to 12%.

of the shoulder[8]. Maurits *et al.* demonstrated that they could separate, almost completely, two types of disorders (myopathies and neuropathies) based on abnormality of ultrasound muscle density and homogeneity[16,17]. In the study by Pillen *et al.*[18], a comparison was made between the sensitivity and specificity of visual *vs* quantitative evaluation of skeletal muscle ultrasound in children suspected of having a neuromuscular disorder (NMD). The quantitative analysis resulted in a higher interobserver agreement (kappa = 0.86) compared with visual evaluation (kappa = 0.53). This indicates that quantification of echo intensity is a more objective and accurate method compared with visual analysis and thus is better suited for the screening task in the diagnostic phase of children with a NMD.

These associations between muscle echogenicity and clinical outcome parameters, as demonstrated in other areas of medicine, may also prove to be useful in pelvic floor research. Measuring echogenicity may add to our understanding of what happens in (sub)total levator ani avulsions or in the recovery process after trauma.

In conclusion, this study showed that 3D/4D ultrasound imaging combined with Matlab software is a reliable method to delineate structures of the pelvic floor in nulliparous women and measure puborectalis muscle echogenicity. Future studies using this parameter may add to our understanding of pelvic floor structural anatomy and function.

REFERENCES

1. Dietz HP. Ultrasound imaging of the pelvic floor. Part II: three-dimensional or volume imaging. *Ultrasound Obstet Gynecol* 2004; **23**: 615–625.
2. Dietz HP, Shek C, Clarke B. Biometry of the pubovisceral muscle and levator hiatus by three-dimensional pelvic floor ultrasound. *Ultrasound Obstet Gynecol* 2005; **25**: 580–585.
3. Dietz HP. Quantification of major morphological abnormalities of the levator ani. *Ultrasound Obstet Gynecol* 2007; **29**: 329–334.
4. Lanzarone V, Dietz HP. Three-dimensional ultrasound imaging of the levator hiatus in late pregnancy and associations with delivery outcomes. *Aust N Z J Obstet Gynaecol* 2007; **47**: 176–180.
5. Shek KL, Kruger J, Dietz HP. The effect of pregnancy on hiatal dimensions and urethral mobility: an observational study. *Int Urogynecol J* 2012; **23**: 1561–1567.
6. Mayans D, Cartwright MS, Walker FO. Neuromuscular ultrasonography: quantifying muscle and nerve measurements. *Phys Med Rehabil Clin N Am* 2012; **23**: 133–148.
7. Pillen S, Verrips A, van Alfen N, Arts IM, Sie LT, Zwarts MJ. Quantitative skeletal muscle ultrasound: diagnostic value in childhood neuromuscular disease. *Neuromuscul Disord* 2007; **17**: 509–516.
8. Tsai YH, Huang KC, Shen SH, Yang TY, Huang TJ, Hsu RWW. Quantification of sonographic echogenicity by the gray-level histogram in patients with supraspinatus tendinopathy. *J Med Ultrasonics* 2013. DOI 10.1007/s10396-013-0516-6
9. Mayans D, Cartwright MS, Walker FO. Neuromuscular ultrasonography: quantifying muscle and nerve measurements. *Phys Rehabil Clin N Am* 2012; **23**: 133–151.
10. Unger CA, Weinstein MM, Pretorius DH. Pelvic floor imaging. *Am J Obstet Gynecol* 2011; **38**: 23–43.
11. Dietz HP, Shek C, De Leon J, Steensma AB. Ballooning of the levator hiatus.*Ultrasound Obstet Gynecol* 2008; **31**: 676–680.
12. Pillen S, van Alfen N, Zwarts M. Ultrasound of muscle. In *Neuromuscular Ultrasound*, Walker FO, Cartwright MS (eds). Elsevier: Philadelphia, PA, 2011; 37–57.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–310.
15. Pillen S, van Dijk JP, Weijers G, Raijmann W, de Korte CL, Zwarts MJ. Quantitative gray-scale analysis in skeletal muscle ultrasound: a comparison study of two ultrasound devices. *Muscle Nerve* 2009; **39**: 781–786.
16. Maurits NM, Bollen AE, Windhausen AEJ, de Jager A, vander Hoeven JH. Muscle ultrasound analysis: normal values and differentiation between myopathies and neuropathies. *Ultrasound Med Biol* 2003; **29**: 215–225.
17. Maurits NH, Beenakker EAC, Van Schaik DEC, Fock JM, Van Der Hoeven JH. Muscle ultrasound in children: normal values and application to neuromuscular disorders. *Ultrasound Med Biol* 2004; **30**: 1017–1027.
18. Pillen S, van Keimpema M, Nievelstein RAJ, Verrips A, van Kruijsbergen-Raijmann W, Zwarts MJ. Skeletal muscle ultrasonography: visual versus quantitative evaluation. *Ultrasound Med Biol* 2006; **32**: 1315–1321.

