VISUAL ASSESSMENT OF SEGMENTAL MUSCLE ULTRASOUND IMAGES IN SPINA BIFIDA APERTA

RICK BRANDSMA,* RENATE J. VERBEEK,* NATASHA M. MAURITS,* JANNEKE T. HAMMINGA,* OEBELE F. BROUWER,* JOHANNES H. VAN DER HOEVEN,* HUIBERT BURGER,† and DEBORAH A. SIVAL†

*Department of Neurology; †Pediatrics; and ‡Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

(Received 27 July 2011; revised 15 March 2012; in final form 10 April 2012)

Abstract—In spina bifida aperta (SBA), spinal MRI provides a surrogate marker to estimate muscle damage caudal to the myelomeningocele (MMC). This muscle damage by the MMC can be quantified by intra-individual comparison of muscle ultrasound density (MUD) cranial versus caudal to the MMC (dMUD = [MUDcranial-to-the-MMC] − [MUDcaudal-to-the-MMC]). Quantitative dMUD assessment requires time, equipment and expertise, whereas it could also be visually determined by differences in muscle echodensity cranial vs. caudal to the MMC (visual-dMUD). If visual and quantitative dMUD correspond, visual dMUD assessment could provide a clinical screening parameter. In 100 SBA muscle ultrasound recordings of patients with various MMC levels, we aimed to compare quantitative dMUD (dMUD = [MUDcalf-muscle/S1] − [MUDquadriceps-muscle/L2-L4]) with visual dMUD assessments by 20 different observers. Results indicate that quantitative dMUD can be visually detected (sensitivity 86%; specificity 57%), implicating that visual dMUD screening could provide a quick, clinical screening tool for muscle impairment by the MMC.

Key Words: Spina bifida, Muscle, Ultrasound, Density, Visual assessment, Myotomes, Myelomeningocele, Child, Neuromuscular damage, Second hit.

INTRODUCTION

In spina bifida aperta (SBA), spinal magnetic resonance imaging (MRI) provides a surrogate marker to estimate muscle impairment caudal to the myelomeningocele (MMC). However, because of abnormal segmental innervation caudal to the MMC, actual myopathic changes appear more directly associated with muscle function loss than the MMC demarcation by MRI (Sival et al. 2004; Verbeek et al. 2009). Such myopathic changes (involving reduction in muscle water content, fat deposition and fibrosis) may induce an increased reflection of the muscle ultrasound beam (Pillen et al. 2009), resulting in increased muscle ultrasound density (MUD) (Maurits et al. 2003, 2004; Pillen et al. 2008a, 2008b; Verbeek et al. 2009). As a consequence of the segmental organization of the spinal cord, SBA myotomes caudal to the MMC will be more severely affected than myotomes cranial to the MMC (Sival et al. 2003). Such intra-individual differences in muscle damage (caudal and cranial to the MMC) are indicated by dMUD (calculated as: dMUD = [MUDcaudal-to-MMC] − [MUDcranial-to-MMC]).

dMUD quantification requires time, equipment and expertise, whereas MMC differences between myotomes cranial and caudal to the MMC could also be visually determined. If quantitative and visual dMUD assessments correspond, visual dMUD assessment would provide a global, fast and easily applicable neuromuscular screening tool.

In perspective of the above, our primary aim was to compare quantitative dMUD (i.e., the golden standard) with visual dMUD assessments by diversely skilled neuropsychiatric examiners. If diverse observers could visually assess quantitative dMUD with acceptable sensitivity, visual dMUD assessment could be applicable for clinical screening purposes. Our secondary aim was to evaluate whether quantitative and visual dMUD outcomes change into the same direction under different MMC conditions. This is based upon the reasoning that altered MMC conditions would be expected to have the same impact upon quantitative- and visual-dMUD outcomes. We, therefore, compared quantitative and visual dMUD between...
different MMC subgroups, involving: (1) MMC levels cranial or caudal to L4; and (2) postnatal age younger or older than 3 months. The first subgroups are based upon present or absent involvement of the quadriceps muscle (influencing dMUD = [MUD\text{calf-muscle/S1}] − [MUD\text{quadriceps-muscle/L2-L4}]) (for characterization of myotomes, see Ropper and Samuels 2009). In accordance with the “second hit” hypothesis, the latter subgroups are based upon the timing of perinatal damage, which is superimposed upon the congenital neural tube defect (involving perinatal mechanical, chemical and vascular spinal damage [Sival et al. 1997, 2004, 2008]). After a postnatal time-interval of 3 months, this “second hit” of neural damage would be expected to induce an increase in MUD outcomes caudal to the MMC. In perspective of the above, we, thus, hypothesized that both MMC levels and postnatal age could influence quantitative and visual dMUD outcomes into the same direction.

In the present SBA study, we thus aimed: (1) to compare quantitative dMUD (i.e., the “golden standard”) with visual dMUD screening results (categorized for different observers with various experience characteristics); and (2) to compare the forthcoming quantitative and visual MUD results for different underlying MMC conditions.

METHODS

Differential muscle ultrasound density

The medical ethics committee of the University Medical Center Groningen, The Netherlands approved the present study. With informed consent by the parents, we retrospectively included all assessable SBA muscle image sets, which were recorded between 2004 and 2010 in 52 SBA children (median gestational age of 6 [range 0–19] months). In these 52 SBA infants, we obtained all (n = 100) SBA muscle image sets of sufficient quality for off-line review (concerning the left and/or right leg). In these 100 MUD data sets, MMC was located: at- or cranial to L1 (n = 2); at L2-4 (n = 45); at L5-S2 (n = 50) and caudal to S2 (n = 3) (i.e., median MMC at L5 [range Th11–S3]). In accordance with previously described methods (Maurits et al. 2003, 2004; Verbeek et al. 2009), we assessed MUD of quadriceps (L2-L4) and calf muscles (S1) (Fig. 1) and quantified dMUD by subtraction. Each SBA MUD set was, thus,
derived from the quadriceps and calf muscles in the same leg in the same SBA infant. All muscle ultrasound recordings were performed with the same ultrasound equipment (General Electric Healthcare logiq 9 ultrasound equipment; General Electric, Jiangsu, China) under standardized conditions (for muscle ultrasound gain, dynamic range, compression and time-gain-compensation). We used a linear transducer (14 MHz; gain of 47 dB) and three focal points. For standardization purposes and reproducibility by others, we deliberately did not use time gain compensation (Pillen et al. 2006). An excess of ultrasound gel prevented skin impression. With respect to standardized reference points, we recorded transverse ultrasound images of the quadriceps muscle (probe placed half-way between trochanter major and lateral knee joint cleft) in supine position and of the calf muscle (probe placed at position of maximum circumference) in prone position, which allowed passive muscle relaxation. For digital quantification (Maurits et al. 2003, 2004), we stored five ultrasound images per muscle and determined MUD of the quadriceps and calf muscle within a well-defined region-of-interest (ROI; excluding the surrounding fascia) (Pillen et al. 2006) by Adobe Photoshop (San Jose, CA, USA) [MUD (average pixel value) ranging from 0 (black)–255 (white)]. The ROI for the quadriceps muscle consisted of the cross-sectional area of the rectus femoris and vastus intermedius and for the calf muscle of the cross-sectional area of the gastrocnemius and soleus muscle (Pillen et al. 2006). After exclusion of the highest and the lowest MUD value, we calculated the mean of the three remaining MUD values to minimize variation.

To control for potential segmental MUD differences in healthy controls (i.e., non-MMC children), we assessed and compared MUD of quadriceps and calf muscles in 13 healthy control children (age matched with the present SBA study population). In healthy control children (median age 6 [range 0–19] months), MUD-calf was similar to MUD-quadriceps muscle (MUD_{calf-muscle} vs MUD_{quadriceps-muscle}: 83 [59–106] vs 76 [55–96]; medians [ranges]; NS). In SBA, we could therefore regard a “positive” dMUD as a pathologic consequence of the MMC (quantified by: dMUD = [MUD_{calf-muscle/S1}] – [MUD_{quadriceps-muscle/L2-L4}]).

**Observers**

For visual dMUD assessment, we recruited 20 observers from the medical staff of the University Medical Center Groningen, The Netherlands. All observers were involved in the assessment and treatment of patients with neuromuscular disorders and/or neurulation defects (including medical specialists, interns, neuropediatric students, technicians and scientists from the departments of pediatrics, neurology, neuropathology, and myology), either with or without experience in myology and/or muscle ultrasound recordings (for characteristics see Table 1). Presence of muscle ultrasound experience was defined by performance and interpretation of muscle ultrasound assessments for more than 3 months (full-time). Presence of myology experience was defined by active and independent clinical participation in the university hospital’s neuromuscular team for more than 3 months (involving neurophysiologists, neuropathologists, neurophysiologists and/or pediatric neurologists). All observers were asked to indicate their own experience (with myology and/or muscle ultrasound) anonymously. We checked whether the anonymously self scored experience levels concurred with the pre-estimated experience levels of the total observer group. The percentage of observers with experience (in myology and/or muscle ultrasound) was 50% with an equal distribution among observers with experience in myology and muscle ultrasound (each 25%). The percentage of observers without experience (neither in muscle ultrasound, nor in myology) was 50%.

Observers were not informed about clinical data of the SBA infant and were excluded from the preparation of the test slides. In addition to a manual with a written explanation, all observers received a power point file containing 100 different SBA muscle ultrasound image sets for visual dMUD assessment. Each muscle ultrasound image set was presented as a separate test slide, with one image of the quadriceps muscle (left side of the slide) and one image of the calf muscle (right side of the slide). In quadriceps and calf muscles, we circumscribed the ROI for quantitative and visual dMUD assessment. To avoid interobserver variation by the delineation of the ROI, we circumscribed the ROI for all observers. Observers were instructed to indicate whether MUD within the ROI of the calf muscle (indicated on the right) was higher than MUD within the ROI of the quadriceps muscle (indicated on the left), or not. Observers were not allowed to consider other muscle characteristics than muscle

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of assessors</th>
<th>Experience in years mean (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Experience in MU and/or MY</td>
<td>10</td>
<td>7.60 (0.25–15)</td>
</tr>
<tr>
<td>a. experience in MU, only</td>
<td>4</td>
<td>3.90 (0.25–15)</td>
</tr>
<tr>
<td>b. experience in MY, only</td>
<td>4</td>
<td>11.60 (3.00–15)</td>
</tr>
<tr>
<td>c. experience in both MU and MY</td>
<td>2</td>
<td>7.25 (0.25–15)</td>
</tr>
<tr>
<td>II. No experience in MU or in MY</td>
<td>10</td>
<td>–</td>
</tr>
</tbody>
</table>

Observers characterized according to self scored experience in muscle ultrasound (MU) and myology (MY). The “experienced” observer group I (n = 10) is subdivided in subgroup Ia (n = 4), Ib (n = 4) and Ic (n = 2).
echogenicity within the indicated ROI. All visual scores were subdivided according to the self scored experience in myology or muscle ultrasound.

Since dMUD has to be substantial enough to allow visual discrimination, we first assessed the smallest dMUD (dMUD cut-off point) at which a quantitative dMUD can be visually recognized with a sensitivity exceeding 80% (i.e., the predefined optimum cut-off point). We, therefore, calculated mean sensitivity and specificity of visual dMUD for incremental dMUD cut-off values (i.e., for 0–5; 5–10; 10–15, etc.).

Secondary aim

For the secondary aim, we categorized all image sets according to MMC levels (cranial and caudal to L4) and postnatal age (0–3 vs >3 months postnatal age). In all infants, the upper border of the MMC was indicated by neonatal spinal MRI. Quantitative- and visual-dMUD outcomes were subsequently compared for the above described MMC conditions.

Statistics

We performed statistical analysis with SPSS 16.0 for Windows (SPSS Inc. Chicago, IL, USA). For the visual recognition of quantitative dMUD outcomes by 20 observers, we determined the sensitivity and specificity for all possible cut-off points. Results were compiled into a receiver operating characteristic (ROC) curve. We subsequently determined the optimal quantitative dMUD cut-off point at which the sensitivity of visual recognition of dMUD exceeded 80%.

Sensitivity and specificity of visual dMUD recognition was normally distributed, as shown by Shapiro Wilk test. We compared sensitivity and specificity of visual dMUD recognition for experienced and inexperienced observer groups by the Student’s t-test. Interobserver agreement was assessed by Cohen’s κ test. An acceptable range for interobserver agreement was pre-defined as 0.40–0.60 (Fleiss 1981; Landis and Koch 1977). We subdivided all muscle ultrasound images in accordance with radiologic MMC levels and age. As shown by the Shapiro Wilk test, MMC levels and age were not normally distributed. We, therefore, assessed and compared (quantitative and visual) dMUD outcomes for MMC levels (cranial or caudal to L4) and for age (younger vs. older than 3 months) by Mann Whitney test. Statistical significance was set at a p < 0.05.

RESULTS

Quantitative vs visual dMUD assessment by three observer groups

Mean sensitivity and specificity for all dMUD cut-off points are shown in Table 2. The ROC curve for visual
dMUD assessment is shown in Figure 2. Visual dMUD recognition discriminated a quantitative dMUD of 10 grey-values with a sensitivity of 86% and specificity of 57%. The percentage of false negatives was 13%. The area-under-the-curve was .871.

The collected outcomes of all self-scored observer-experience levels resembled pre-estimated experience levels of the whole group (50% had experience in muscle ultrasound and/or myology; 50% had no experience), for further subdivision into observer categories, see Table 1. Comparing visual dMUD outcomes between observer categories revealed a lower sensitivity for myology-experienced than for myology-inexperienced groups.

Table 2. Sensitivity and specificity of dMUD cut-off points

<table>
<thead>
<tr>
<th>Cut-off points (in grey values)</th>
<th>Mean sensitivity (in %)</th>
<th>Mean specificity (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0</td>
<td>74.3</td>
<td>87.7</td>
</tr>
<tr>
<td>&gt;5</td>
<td>80.4</td>
<td>76.0</td>
</tr>
<tr>
<td>&gt;10</td>
<td>86.1</td>
<td>56.9</td>
</tr>
<tr>
<td>&gt;15</td>
<td>91.5</td>
<td>54.9</td>
</tr>
<tr>
<td>&gt;20</td>
<td>96.6</td>
<td>45.6</td>
</tr>
<tr>
<td>&gt;25</td>
<td>98.6</td>
<td>42.0</td>
</tr>
<tr>
<td>&gt;30</td>
<td>98.6</td>
<td>36.9</td>
</tr>
<tr>
<td>&gt;35</td>
<td>100</td>
<td>35.9</td>
</tr>
<tr>
<td>&gt;40</td>
<td>100</td>
<td>35.2</td>
</tr>
<tr>
<td>&gt;45</td>
<td>100</td>
<td>33.6</td>
</tr>
</tbody>
</table>

Visual dMUD scores by 20 assessors. Cut-off points indicate mean sensitivity and specificity for incremental dMUD values (ranging from 0–45). dMUD = [MUDcalf-muscle] – [MUDquadriceps-muscle].

Fig. 2. ROC-curve of visual dMUD assessment. The x-axis indicates 1-specificity (false positive rate) for the visual dMUD recognition. The y-axis indicates the sensitivity for the visual dMUD recognition. The inserted dots (in the curve) represent a dMUD cut-off point of 5, 10 and 15 grey values. At the cut-off point of 10 grey values, sensitivity of 86% has exceeded the 80% level. The area under the curve for visual dMUD recognition is .871. ROC curve = receiver operating characteristic curve; dMUD = difference in muscle ultrasound density.
In pediatric studies, quantitative MUD assessment has been described as a reliable, noninvasive technique for the detection of neuromuscular disease (Maurits et al. 2003, 2004; Pillen et al. 2003, 2008a, 2008b). However, quantitative MUD assessment requires time, equipment and expertise, which may limit general application. In this perspective, it was hypothesized that visual MUD assessment could provide an easier assessable screening tool (Heckmatt et al. 1982; Pillen et al. 2006; Zuberi et al. 1999). Accordingly, it was previously shown that visual muscle ultrasound assessment could discern Heckmatt’s scores but with a lower sensitivity than quantitative MUD assessment (71% and 87%, for visual and quantitative assessment, respectively [Pillen et al. 2006]). In the present SBA MUD study, we, therefore, applied a more simplified visual screening method involving discrimination between intra-individual MUD differences (dMUD). Results indicated that a MUD difference of at least 10 grey-values could be visually discerned with a sensitivity exceeding 80%. Applying this optimum dMUD cut-off point (of 10 grey-values) for visual MUD outcomes, revealed an acceptable accuracy (81%) and sensitivity (86%) of dMUD detection, whereas specificity (57%) appeared low. These characteristics are typical for a “rule out test,” implicating that if a positive quantitative dMUD is present, it can be visually discerned in 86% of the cases.

Conversely, a negative visual dMUD screening outcome would make a positive quantitative dMUD outcome unlikely (i.e., false negative rate is 13%). In SBA children, this implicates that intra-individual visual dMUD assessment could provide a quick and noninvasive screening method for segmental muscle damage. However, in case of an unexpected visual dMUD outcome, quantitative dMUD confirmation (i.e., golden standard) may be advisable.

Interestingly, inexperienced and muscle ultrasound experienced observers achieved higher sensitivities than myology-experienced observers. Since the self scored and pre-estimated experience characteristics did not

### Table 3. Sensitivity and specificity of visual dMUD recognition

<table>
<thead>
<tr>
<th>Experience</th>
<th>Total group</th>
<th>Myology</th>
<th>Muscle ultrasound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All N = 20</td>
<td>Yes N = 10</td>
<td>No N = 10</td>
</tr>
<tr>
<td></td>
<td>t-test*</td>
<td>t-test*</td>
<td>t-test*</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>86.3%</td>
<td>83.3%</td>
<td>89.3%</td>
</tr>
<tr>
<td>Specificity</td>
<td>59.6%</td>
<td>62.4%</td>
<td>56.9%</td>
</tr>
</tbody>
</table>

Sensitivity and specificity of visual dMUD recognition (at the optimal cut-off point of 10 grey-values) according to the observers’ experience. There were no significant differences between the combined experienced and inexperienced observer groups. Observers who were more experienced in

\( p = 0.012 \), see Table 3. At a cut-off point of 10 grey values, interobserver agreement was 0.514. Subdivision in inexperienced, muscle ultrasound-experienced and myology-experienced observer subgroups revealed similar kappa scores (0.536, 0.541 and 0.459 respectively [NS]), which were all three within the pre-defined acceptable range for \( \kappa \) scores (0.40–0.60).

The relation between dMUD and various MMC conditions (MMC level and age)

Associating quantitative dMUD with MMC levels revealed higher dMUD outcomes for MMC levels caudal than cranial to L4 \( p = 0.032 \). The sensitivity of visual dMUD recognition (i.e., percentages of assessors discerning a positive dMUD) revealed also higher outcomes for MMC levels caudal than cranial to L4 \( p = 0.005 \).

Associating postnatal age subgroups (younger vs. older than 3 months of age) with included MMC levels, did not reveal significant differences \( p = 0.857 \). Quantitative dMUD assessment revealed higher outcomes after- than before- the third month of postnatal life \( p = .008 \). Accordingly, sensitivity of visual dMUD recognition (i.e., percentages of assessors discerning a positive dMUD) was also higher after than before the third month of postnatal life \( p = 0.027 \).

### DISCUSSION

In children with SBA, we aimed to associate intra-individual visual and quantitative dMUD outcomes. Independent of the observer characteristics, results indicated a positive relationship between visual and quantitative dMUD outcomes. Furthermore, both quantitative and visual dMUD parameters were changed into the same direction by different MMC conditions (MMC level and postnatal age). Present SBA data may implicate that visual dMUD could provide a simple and quick screening method to estimate potential muscle damage caudal to the MMC.
differ, we can not attribute this finding to falsely self-scored experience levels. This may implicate that a simplified test (involving discrimination between grey-values within a predefined ROI) rather than a complex test (such as qualitative visual interpretation involving Heckmatt scale), could exert a more favorable effect upon the outcomes by the inexperienced than by myology-experienced observers. Since the ROI was clearly indicated to all observers and since the observers were not allowed to evaluate any other aspects than MUD within the ROI, these outcomes do not necessarily implicate that inexperienced observers are technically more capable of visual dMUD screening. Especially when visual MUD assessment would be applied for more complex and diverse neuromuscular diseases (requiring qualitative interpretation of the image), we would expect that both myology- and muscle ultrasound-experienced observers would achieve better results than inexperienced observers.

Finally, we observed that both quantitative and visual dMUD screening outcomes change into the same direction according to the segmental MMC location and postnatal age of the child (i.e., the time period to induce maximal histologic muscle alterations). The latter age-dependent finding appears reflective of the consequences by the second hit hypothesis, involving ongoing perinatal spinal trauma (inducing secondary muscle denervation damage and subsequently delayed muscle fibrosis and muscle fat deposition, thereafter [Sival et al. 2008]).

All together, in children with SBA we conclude that visual dMUD assessment could provide a simple screening method for a quick estimation of segmental muscle damage.

Acknowledgments—The authors thank all observers for their participation and Mrs. T. Bijzitter, Mrs. Bijmolt and Mrs. H. Kunst for their excellent administrative help. The authors do not declare any conflicts of interest.

REFERENCES


