A novel segmentation framework dedicated to the follow-up of fat infiltration in individual muscles of patients with neuromuscular disorders

Augustin C. Ogier, Linda Heskamp, Constance P. Michel, Alexandre Fouré, Marc-Emmanuel Bellemare, Arnaud Troter, Arend Heerschap, David Bendahan

A novel segmentation framework dedicated to the follow-up of fat infiltration in individual muscles of patients with neuromuscular disorders

Augustin C. Ogier[1,2] | **Linda Heskamp**[3] | **Constance P. Michel**[2] | **Alexandre Fouré**[2,4] | **Marc-Emmanuel Bellemare**[1] | **Arnaud Le Troter**[2] | **Arend Heerschap**[3] | **David Bendahan**[2]

[1]Aix Marseille University, Université de Toulon, CNRS, LIS, Marseille, France

[2]Aix Marseille University, CNRS, CRMBM, Marseille, France

[3]Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, Netherlands

[4]Laboratoire Interuniversitaire de Biologie de la Motricité, Université Claude Bernard Lyon 1, Villeurbanne, France

Correspondence
Augustin C. Ogier, Laboratoire d'Informatique et Systèmes, Aix-Marseille Université – St. Jérôme, Avenue Escadrille Normandie Niemen, Marseille, Provence-Alpes-Côte d'Azur 13013, France.
Email: augustin.ogier@gmail.com

Funding information
European Union Seventh Framework Program on DM1 (OPTIMISTIC); Grant/Award No. 305697.

Purpose: To propose a novel segmentation framework that is dedicated to the follow-up of fat infiltration in individual muscles of patients with neuromuscular disorders.

Methods: We designed a semi-automatic segmentation pipeline of individual leg muscles in MR images based on automatic propagation through nonlinear registrations of initial delineation in a minimal number of MR slices. This approach has been validated for the segmentation of individual muscles from MRI data sets, acquired over a 10-month period, from thighs and legs in 10 patients with muscular dystrophy. The robustness of the framework was evaluated using conventional metrics related to muscle volume and clinical metrics related to fat infiltration.

Results: High accuracy of the semi-automatic segmentation (mean Dice similarity coefficient higher than 0.89) was reported. The provided method has excellent reliability regarding the reproducibility of the fat fraction estimation, with an average intraclass correlation coefficient score of 0.99. Furthermore, the present segmentation framework was determined to be more reliable than the intra-expert performance, which had an average intraclass correlation coefficient of 0.93.

Conclusion: The proposed framework of segmentation can successfully provide an effective and reliable tool for accurate follow-up of any MRI biomarkers in neuromuscular disorders. This method could assist the quantitative assessment of muscular changes occurring in such diseases.

KEYWORDS
image registration, image segmentation, muscle, neuromuscular disorders

1 | INTRODUCTION

Patients with neuromuscular diseases experience progressive loss of muscle strength. Histopathological analysis of muscle biopsies often reveals signs of fatty infiltration. Quantification of the corresponding pathological process is of great value for understanding natural history of a given disease[1-3] and for reliable evaluation of therapeutic strategies.[4,5]

Magnetic resonance imaging has become a tool of choice for the investigation of neuromuscular diseases, thanks to its noninvasive nature and its ability to distinguish fat and muscle tissue.[6] As such, MRI allows quantitative assessment of the amount of fat infiltration by determining the fraction of fat within the muscle. To do so, muscles of interest have to be delineated. The corresponding segmentation process has primarily been performed manually, making the approach operator-dependent and time-consuming.[7] As a result, often only a limited number of slices have been selected.[3,8] Considering that fat replacement is nonuniformly distributed along the proximo-distal axis in muscles,[1,9] a single slice or a limited number of slices might not accurately represent the entire muscle. In addition, each individual muscle can be affected differently and varies among patients and neuromuscular disorders.[10-12] Therefore, it is of utmost importance to assess muscles individually, instead of muscle groups. Reliable automatic segmentation methods are still warranted for individual muscles.

Several semi-automated and automated methods have been reported in the literature with the aim of segmenting muscle groups or individual muscles in MR images of healthy volunteers.[13-18] Although promising, none of these methods has been tested for MR images recorded in patients with neuromuscular disorders. Automatic segmentation methods have been reported for images of obese patients,[19,20] but only for the discrimination between the muscle as a whole and the adipose tissue. More recently, a few studies have addressed the issue of distinguishing muscle, intramuscular adipose tissue, and subcutaneous adipose tissue in patients with a severe fat infiltration.[21-24] These approaches have considered the muscle compartment as a whole or as muscle groups, which might not be optimal in neuromuscular disorders in which individual muscles will be affected differently.[11,12] Segmentation of individual muscles has only been performed in healthy subjects, in whom fat infiltration and atrophy are not confounding factors for the automatic muscle segmentation.

We previously combined 2D nonlinear registration methods with the aim of retrieving the whole 3D muscle segmentation on the basis of a propagation of 2D manually segmented masks for a limited number slices. This approach reported promising results for the segmentation of the 4 quadriceps femoris muscles from MR images recorded in healthy subjects.[25] As an extension of this seminal work, the first purpose of the present study was to assess the performance of the corresponding method for the segmentation of all individual leg muscles in MR images recorded in patients with neuromuscular disorders.

Furthermore, quantification of muscle volumes and fat infiltration in individual muscles is also of interest for longitudinal assessments such as natural history studies or clinical trials. In that respect, segmentation of individual muscles has to be repeated multiple times along the assessment period. Therefore, the second aim of this study was to further extend our initial work toward a longitudinal aspect. More specifically, we proposed a new propagation tool based on a 3D nonlinear registration method that is dedicated to follow-up studies. In that respect, we assessed the corresponding performance regarding changes over repeated sessions with an initial delineation only performed at baseline. The proposed segmentation framework has been designed to save time without compromising the accuracy of the whole 3D muscle segmentation for multiple data sets recorded longitudinally. The robustness of the framework was evaluated using conventional metrics related to muscle volume and clinical metrics related to fat infiltration.

2 | METHODS

2.1 | Subjects

We randomly selected 12 patients (5 women) with genetically confirmed myotonic dystrophy type 1 (DM1), who participated in the randomized controlled trial OPTIMISTIC.[26] They were 46 ± 12 years old, had a disease severity score of 3.0 ± 0.8 (MIRS score[27]) and could walk on average 411 ± 90 m during a 6-minute walk test.

2.2 | Magnetic resonance imaging protocol

The thigh and lower leg of the DM1 patients were examined twice, 10 months apart, with a 3T MR system (TIM Trio, Siemens, Erlangen, Germany) using a spine/phased array coil combination. We acquired a 3D Dixon sequence (TR = 10 ms, flip angle = 3°, in-plane FOV = 256 × 192 mm^2, matrix = 256 × 192, number of slices = 32, slice thickness = 5 mm) with either a 2-point Dixon (TE1/TE2 = 2.45/3.675 ms) or a 3-point Dixon (TE1/TE2/TE3 = 2.31/3.68/5.07 ms). The same protocol was used for the thigh and leg and ensured that baseline and follow-up scans were acquired with the same version of the Dixon protocol (i.e., or both time points with a 2-point version or both time points with a 3-point version). Fat (F) and water (W) images were generated using the manufacturer's reconstruction, and a fat fraction (FF) map was calculated voxel-wise as F/(F+W). Furthermore, 4 subjects were scanned twice on the same day in order to assess intrasubject repeatability. The study was approved by the local human research committee and was conducted in conformity with the Declaration of Helsinki (version October 2013) and the Medical Research Involving Human Subjects Act. Prior written informed consent was obtained from all subjects.

2.3 | Ground truth segmentation

An expert (L.H.) manually segmented each individual muscle of the thigh and lower leg on the 25 middle slices of the

out-of-phase images using FSLview, the 3D viewer included in the FSL toolbox.[28] Individual muscles were delineated on all 25 slices approximately 1 voxel away from the subcutaneous fat, fasciae, and large blood vessels, to avoid contaminations that could bias the FF values. The corresponding masks were considered as the ground truth and were used to assess the performance of the semi-automatic segmentation method. The following muscles were delineated: adductor longus, adductor magnus, long head of the biceps femoris, short head of the biceps femoris (BFS), gracilis, rectus femoris (RF), sartorius, semimembranosus, semitendinosus, vastus intermedius, vastus lateralis, vastus medialis (VM) and extensor digitorum/hallucis longus, fibularis brevis/longus, gastrocnemius lateralis (GL), gastrocnemius medialis (GM), soleus, tibialis anterior, and tibialis posterior (TP) for the thigh and the lower leg, respectively (Figure 1). The time for the manual segmentation of a given data set was approximately 2 hours for the lower leg and 3 hours for the thigh.

2.4 | Segmentation propagation

Our semi-automatic segmentation framework is outlined in Figure 2. Two different successive propagation processes were involved: 1 for transversal propagation and 1 for longitudinal propagation for the segmentation of baseline and follow-up data sets, respectively. Segmentation propagations were performed using the 4 Dixon contrast independently, to determine the optimal contrast for the propagation processes.

2.4.1 | Transversal propagation

On the baseline data set, we selected the first and the last slice of the manual gold-standard segmentation defining the 3D region of interest together with additional slices for which a muscle was appearing or disappearing (step 1 in Figure 2). For the thigh, the manual segmentation of 3 supplementary slices were required in order to take into account

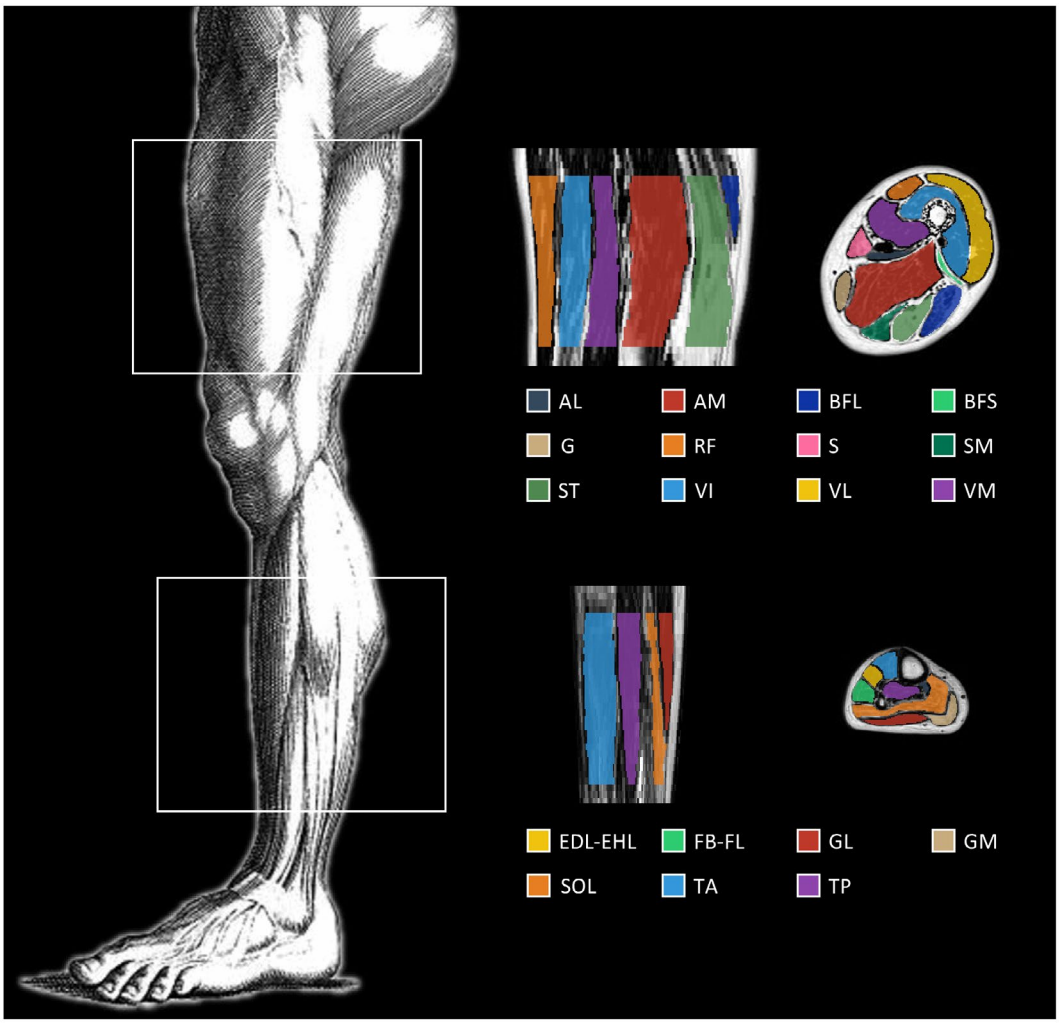

FIGURE 1 Overview of the segmented muscles in the thigh and lower leg. Abbreviations: AL, adductor longus; AM, adductor magnus; BFL, long head of the biceps femoris; BFS, short head of the biceps femoris; EDL-EHL, extensor digitorum/hallucis longus; FB-FL, fibularis brevis/longus; G, gracilis; GL, gastrocnemius lateralis; GM, gastrocnemius medialis; RF, rectus femoris; S, sartorius; SM, semimembranosus; SOL, soleus; ST, semitendinosus; TA, tibialis anterior; TP, tibialis posterior; VI, vastus intermedius; VL, vastus lateralis; VM, vastus medialis

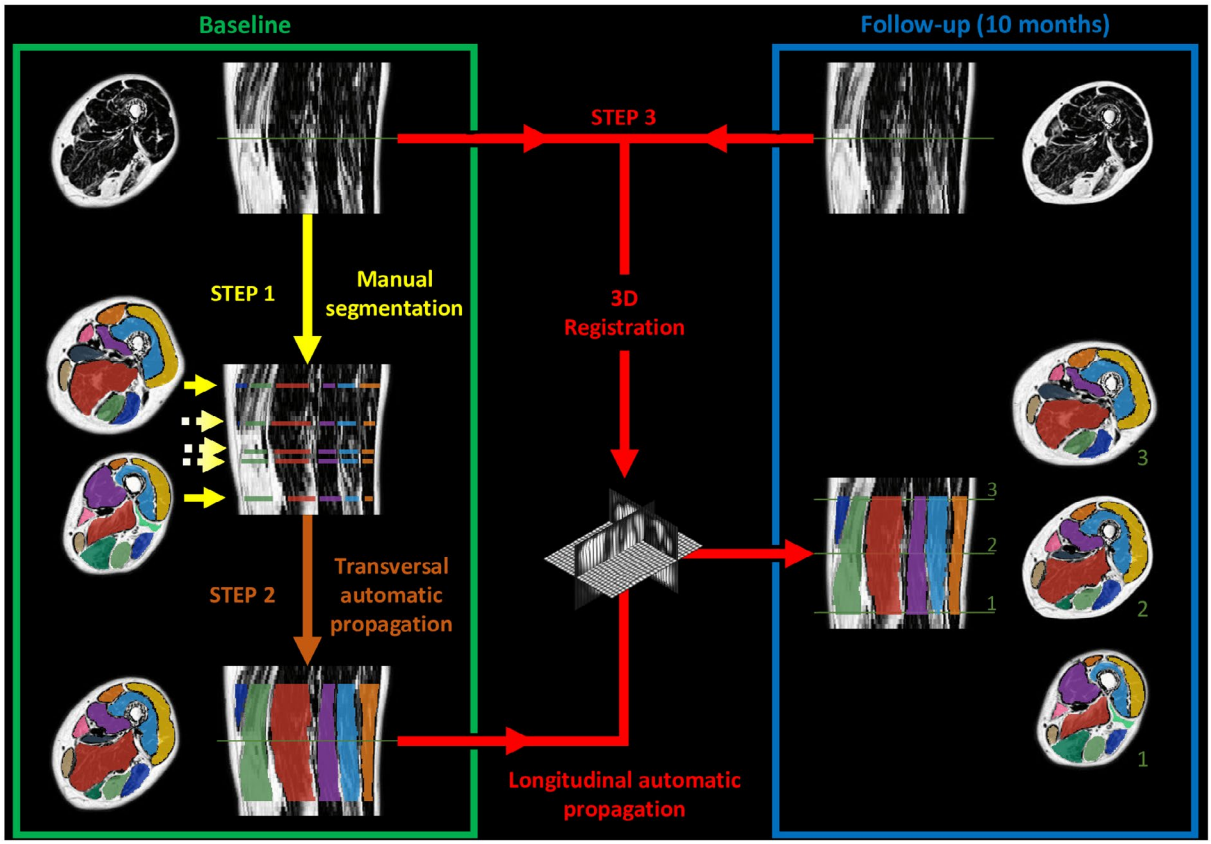

FIGURE 2 Schematic overview of the semi-automated segmentation pipeline. The manual delineation of individual muscles was initially performed for a few axial slices of each baseline data set (step 1). The corresponding masks were then transversally propagated to the whole 3D region of interest (step 2). Subsequently, a 3D nonlinear registration step was performed between baseline and follow-up data sets so that a longitudinal propagation of the baseline 3D segmentations was possible (step 3)

the emergence of both RF and adductor longus muscles and the termination of the BFS muscle. For the lower leg, the manual segmentation of 2 supplementary slices were needed for the emergence of gastrocnemii muscles. These nine 2D manual segmentations were then transversally propagated to the 41 remaining slices of the 3D region of interest using several 2D nonlinear registration approaches as we previously described[25] (step 2). For this transversal propagation aspect, the method took advantage of the shape information from the initial manual segmentations, with no other previous information regarding muscles shapes, together with the grayscale anatomical information provided by MR images, to follow the anatomical variations of muscles along the leg. The propagation process was performed simultaneously on all muscles, so overlapping classes did not occur.

2.4.2 | Longitudinal propagation

The longitudinal propagation was based on a robust registration process performed between the baseline and follow-up data sets. As a prerequisite, a N4 bias correction algorithm[29] was used on both data sets. Then, an initial translation was performed using the centers of mass, and the corresponding result was used to design an optimal affine transformation. Thereafter, a 3D nonlinear B-Spline SyN registration (B-spline symmetric diffeomorphic normalization implemented in the ANTs library[30]) was performed between the pre-aligned baseline and follow-up Dixon data sets. The choice of such a nonlinear diffeomorphic model was motivated by the need to register muscles that might show large deformations between 2 acquisition sessions. The following variables were used for the optimization of the B-spline SyN stage: gradient step = 0.1, cost function as cross correlation, 4 multiresolution steps with, respectively, $100 \times 70 \times 50 \times 10$ iterations max per level, shrink factor = $8 \times 4 \times 2 \times 1$, and smoothing sigmas = $3 \times 2 \times 1 \times 0$ mm. Finally, we applied the B-spline SyN deformation field and the affine transform file on the baseline segmentation to obtain the follow-up segmentation masks. Nearest-neighbor interpolation was applied to each baseline mask to keep the integer values from the original labels. To make sure that, on a subject basis, volumes of interest were comparable for successive acquisition time points, translations along the proximo-distal axis were not allowed for the longitudinal

registration processes. This way, the volumes of interest for both baseline and follow-up MR acquisitions always covered the same area of 25 slices.

2.4.3 | Validation metrics

To assay the performance of our semi-automatic segmentation framework regarding both transversal and longitudinal propagations, metrics related to muscle volume and fat infiltration were used. The manually segmented slices used as input for the transversal propagation process have not been considered in the metric computations.

To assess the similarity between volumes of interest, the Dice similarity coefficient[31] (DSC) was calculated for each muscle. The DSC values can range from 0 to 1 and describe the similarity between the muscle segmentation obtained using our semi-automatic methods (A) and with the manual delineation process (M) (Equation 1). A high DSC value illustrates a high similarity.

$$DSC(A,M) = \frac{2 \cdot |A \cap M|}{|A| + |M|} \quad (1)$$

Regarding the fat infiltration metric, FF values were averaged within the whole set of voxels of the FF map within the regions of interest for both the manual (FF_M) and automatic segmentation (FF_A). The values of FF_M and FF_A were computed separately for both baseline and follow-up data sets.

In addition, we performed 2 types of repeatability assessments to test the repeatability of the fat infiltration quantification through the segmentation and between repeated acquisitions. First, the measurement repeatability was assessed by scanning 4 subjects twice over a single day. In that case, segmentations were performed manually twice (FF_{MR1} and FF_{MR2}) by the same expert. Second, we assessed the interexpert repeatability of the method. Two experts performed the initialization work (i.e., the manual segmentation of a few chosen slices for 4 MRI data sets [thigh and leg] randomly extracted from the baseline data sets). The semi-automated propagation generated 2 sets, A_1 (expert 1) and A_2 (expert 2), from which FF values (FF_{A1} and FF_{A2}) could be computed and compared to the results of the ground-truth manual segmentation (FF_M) of expert 1.

The obtained FF values were compared using the intraclass correlation coefficient (ICC) with a 2, 1 formula[32] and standard error of measurement (SEM). Furthermore, Bland-Altman analysis[33] was performed to analyze the agreement between the manual and automatic methods.

3 | RESULTS

The proposed method was used for the semi-automatic segmentation of 20 data sets from 12 subjects (10 thigh and 10 lower-leg data sets). Two data sets were excluded due to acquisition issues.

3.1 | Semi-automatic propagation compared with manual ground truth

3.1.1 | Dice similarity index

As illustrated in Table 1, DSC values related to different contrasts did not differ for both the transversal and longitudinal propagations. For the sake of clarity, all results shown from here on were obtained by propagating on the water contrast, which provided the highest mean DSC values. Similar results were obtained with the other contrasts.

Figure 3 shows a few examples of the muscle-region segmentation propagation for the thigh and lower leg. For the sake of simplicity, only the most middle slice of the segmented region of interest is shown. There is a large variation in muscle volume, muscle fat infiltration, and subcutaneous fat thickness among subjects. Furthermore, these images represent the variation observed in our data set between a baseline scan and a follow-up. For example, the lower leg shape in patient P10 was clearly different between baseline and follow-up, probably due to different positioning of the leg and surface coils, affecting the longitudinal segmentation. Considering the whole set of muscles, the overall mean DSC value was higher than 0.89, with an average of 91.2 ± 4.3 for baseline data sets (transversal propagation) and 87.9 ± 7.7 for follow-up (longitudinal propagation) data sets (Table 1). The DSC values obtained for the individual muscles ranged from 0.83 to 0.96 for the baseline and from 0.76 to 0.94 for the follow-up data sets (Figure 4). The transversal propagation showed rather homogeneous results with DSC values higher than 0.90 for all muscles but adductor longus, BFS, extensor digitorum/hallucis longus, and GL. The GL had the lowest overall scores for almost every patient. The longitudinal propagation generated more outliers with a particularly low

TABLE 1 Global mean DSC between manual reference and automatic propagation (transversal and longitudinal) segmentation using each Dixon contrast

Dixon image	In phase	Out phase	Water	Fat
$DSC_{Transversal\ Propagation}$	89.5 ± 5.2	90.3 ± 4.8	91.2 ± 4.3	90.9 ± 4.5
$DSC_{Longitudinal\ Propagation}$	80.0 ± 17.0	85.4 ± 13.3	87.9 ± 7.7	87.3 ± 9.1

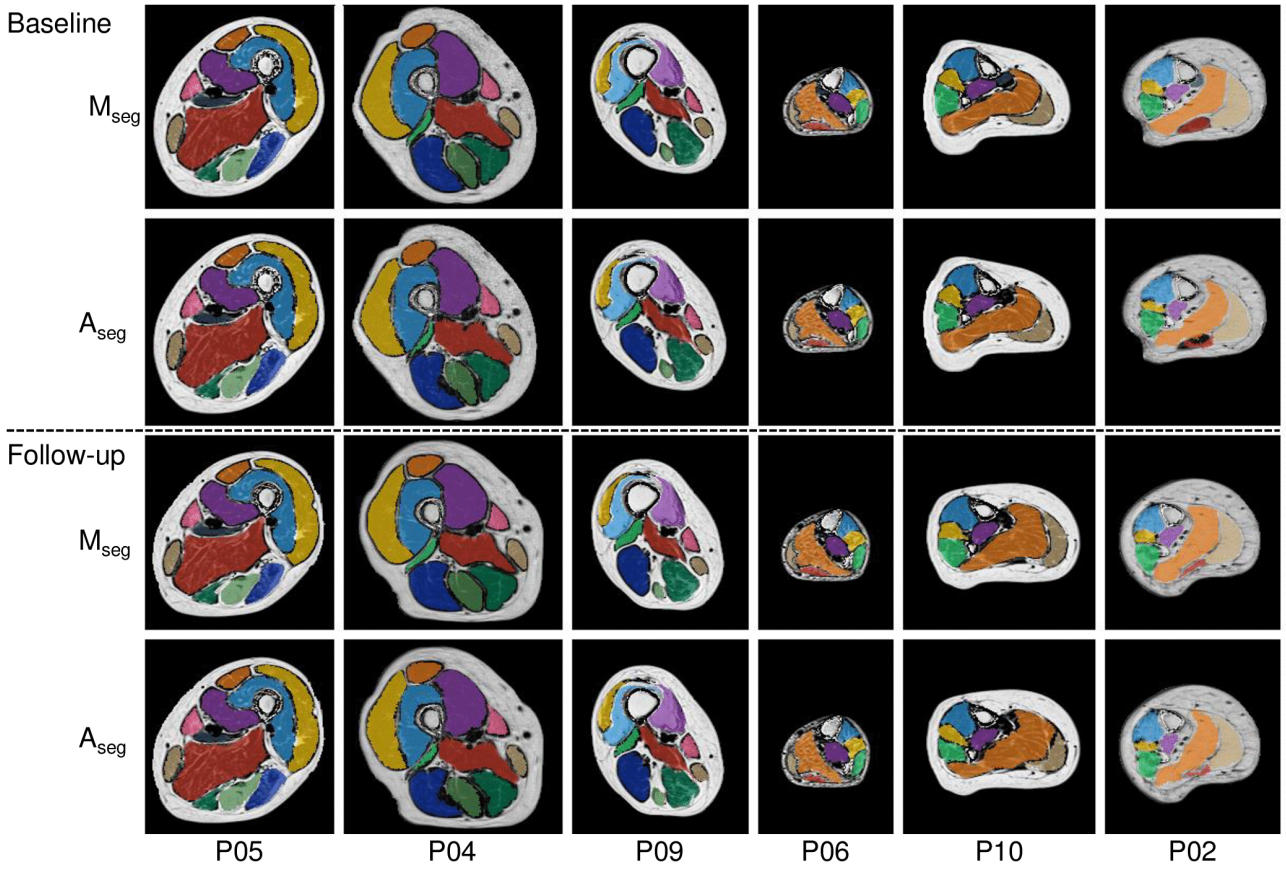

FIGURE 3 Examples of the automatic muscle segmentation for the baseline propagation and the longitudinal follow-up propagation (A_{seg}) compared with the manual segmentation (M_{seg})

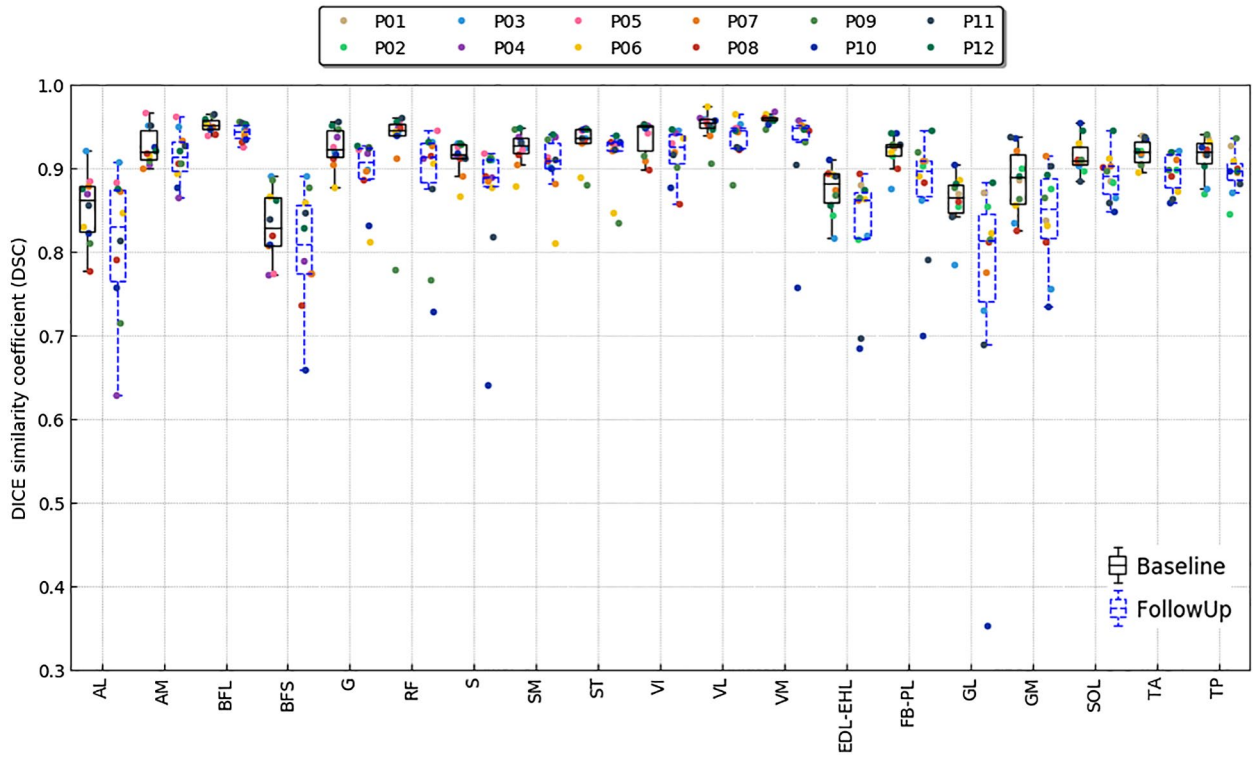

FIGURE 4 Box plots showing Dice individual similarity coefficients (DSC) for each muscle (x-axis). The DSC values were calculated from a comparative analysis between manual and automatic segmentations for both baseline and follow-up data sets. Circles represent DSC values for each subject

DSC value (0.35) regarding the GL muscle in patient P10 (Figure 4).

3.1.2 | Fat fraction

The averaged FF values quantified from the manual and automatically obtained segmentations and the corresponding ICC and SEM values are gathered in Table 2. For the thigh muscles at baseline, FF_M values ranged from 7.8 ± 7.0 % (RF) to 22.1% ± 21.1% (semimembranosus), while FF_A values were very similar and ranged from 8.3% ± 7.8% (RF) to 22.3% ± 21.0% (semimembranosus). For follow-up data sets, FF values were slightly higher with FF_M values in the thigh ranging from 9.5% ± 7.7% (RF) to 24.2% ± 25.9% (semitendinosus), while FF_A values ranged from 11.2% ± 10.5% (RF) to 24.2% ± 25.7% (semitendinosus). In the lower leg, more severe fat infiltration is observed (Table 2). For the baseline data sets, FF_M values in the lower leg ranged from 13.1% ± 17.9% (TP) to 36.8% ± 25.8% (SOL), while FF_A values were very similar and ranged from 13.1% ± 18.1% (TP) to 37.0% ± 27.8% (GM). Similar to what we observed in the thigh muscles, the averaged FF values were slightly higher for the follow-up data sets, with FF_M values ranging from 14.0% ± 17.6% (TP) to 39.3% ± 26.9% (semitendinosus), while FF_A values ranged from 14% ± 17.6% (TP) to 39.9% ± 23.6% (SOL).

3.1.3 | Intraclass correlation coefficient

The corresponding ICC values were very high, with an average score of 0.99, and most values were equal to 1.00. In the baseline data sets, they ranged from 0.98 to 1.00 and 0.99 to 1.00 in the thigh and lower leg, respectively. For the follow-up data sets, ICC values ranged from 0.94 to 1.00 for the thigh muscles and from 0.97 to 1.00 for the lower leg muscles. For each muscle, ICC values were always lower in the follow-up data sets than in the baseline data sets (Table 2), which can be explained by the longitudinal propagation of segmentation errors created in the transversal propagation process.

3.1.4 | Bland-Altman analysis

The Bland-Altman analysis revealed for the FF values quantified from automatic and manual segmentations at baseline a bias of −0.3 percentage point of FF value with 95% limits of agreement ranging between −3.0 and 2.4 points (Figure 5A).

TABLE 2 Mean FF values, intraclass correlation coefficient, and standard error of the measurement measured for the baseline and follow-up data sets

	Baseline				Follow-up (10 months)			
Muscle	FF_M (%)	FF_A (%)	ICC	SEM (%)	FF_M (%)	FF_A (%)	ICC	SEM (%)
AL	14.5 ± 27.3	14.3 ± 26.5	1.00	0.6	15.0 ± 26.8	15.8 ± 26.5	1.00	1.3
AM	12.0 ± 10.7	14.2 ± 10.7	0.98	1.7	12.9 ± 12.0	14.8 ± 11.8	0.99	1.2
BFL	13.1 ± 13.5	13.2 ± 13.7	1.00	0.1	13.9 ± 13.4	14.1 ± 13.5	1.00	0.2
BFS	19.4 ± 23.1	19.6 ± 23.3	1.00	0.3	20.3 ± 22.2	20.6 ± 22.4	1.00	0.4
G	9.7 ± 5.2	9.9 ± 5.3	0.98	0.7	10.6 ± 5.1	11.2 ± 5.7	0.94	1.3
RF	7.8 ± 7.0	8.3 ± 7.8	0.99	0.7	9.5 ± 7.7	11.2 ± 10.5	0.94	2.2
S	16.5 ± 10.0	16.3 ± 9.9	1.00	0.5	17.4 ± 9.5	18.2 ± 9.6	0.98	1.3
SM	22.1 ± 21.1	22.3 ± 21.0	1.00	0.4	23.1 ± 21.0	23.3 ± 21.3	1.00	0.4
ST	22.0 ± 23.5	22.2 ± 23.7	1.00	0.7	24.2 ± 25.9	24.2 ± 25.7	1.00	0.3
VI	17.7 ± 23.4	18.1 ± 22.9	1.00	0.8	19.4 ± 23.7	19.3 ± 23.2	1.00	0.8
VL	11.1 ± 8.9	11.1 ± 8.9	1.00	0.1	11.8 ± 9.5	12.2 ± 9.8	1.00	0.4
VM	13.2 ± 19.6	13.3 ± 19.8	1.00	0.1	14.3 ± 20.6	14.2 ± 20.4	1.00	0.2
EDL-EHL	25.3 ± 18.8	25.4 ± 18.7	1.00	0.3	26.8 ± 19.3	27.2 ± 18.6	1.00	0.8
FB-FL	28.0 ± 21.7	28.1 ± 21.6	1.00	0.3	29.7 ± 21.2	29.7 ± 21.1	1.00	0.4
GL	19.2 ± 23.4	21.7 ± 22.2	0.99	2.7	20.3 ± 23.7	23.4 ± 21.8	0.97	3.7
GM	36.7 ± 28.1	37.0 ± 27.8	1.00	0.9	39.3 ± 26.9	39.5 ± 26.4	1.00	1.2
SOL	36.8 ± 25.8	36.4 ± 25.5	1.00	0.3	40.2 ± 23.8	39.9 ± 23.6	1.00	0.4
TA	24.9 ± 19.9	24.9 ± 19.9	1.00	0.4	26.4 ± 19.6	26.5 ± 19.2	1.00	0.8
TP	13.1 ± 17.9	13.1 ± 18.1	1.00	0.3	13.7 ± 17.6	14.0 ± 17.6	1.00	0.3

Note: Mean FFs were quantified on masks resulting from the automatic (FF_A) and manual (FF_M) segmentations for each muscle. ICC and SEM values were computed for each muscle from a comparative analysis between both methods.

Abbreviations: ICC, intraclass correlation coefficient; SEM, standard error of the measurement.

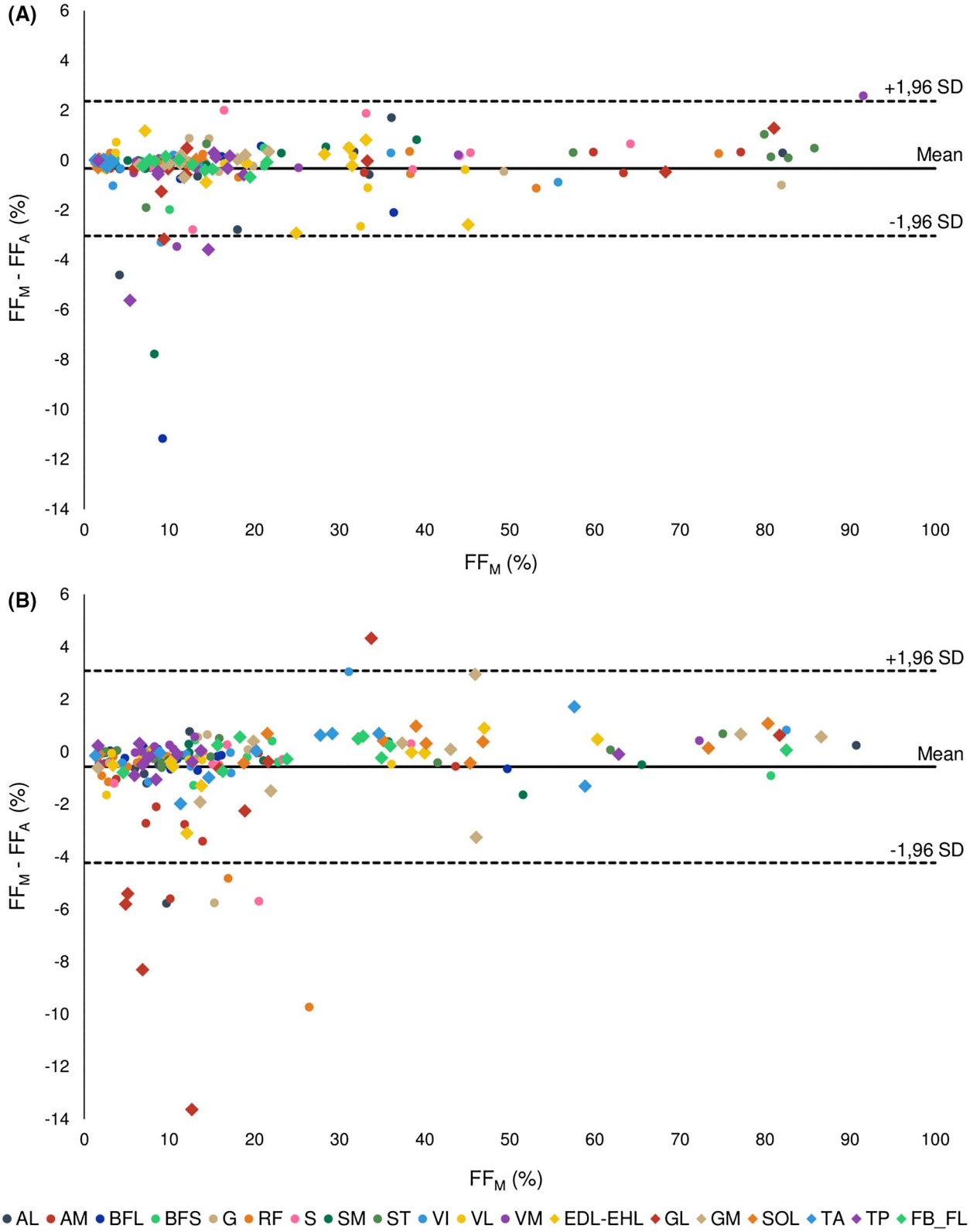

FIGURE 5 Bland-Altman analysis of fat fraction (FF) values quantified for each individual muscle and each patient for the manual (FF_M) and automatic (FF_A) segmentations. A, Baseline data. B, Follow-up data. Central lines indicate the mean difference, and the dashed lines indicate the limits of agreement

The FF values computed from automatic and manual segmentations for follow-up data sets illustrated a bias of −0.5 point with 95% limits of agreement ranging between −4.2 and 3.1 points (Figure 5B). For both the baseline and follow-up data sets, the bias was not related to the muscle's FF (Figure 5).

3.2 | Measurement repeatability

The average FF values and the corresponding ICC and SEM values for the 4 volunteers who were scanned twice on the same day are found in Table 3. In the thigh, FF_{MR1} values ranged from 5.3% ± 2.9% (VM) to 11.3% ± 8.4% (BFS), while FF_{MR2} values were quite similar and ranged from 4.8% ± 2.6% (adductor magnus) to 11.3% ± 8.1% (sartorius). Similar to the FF values in Table 2, the fat infiltration was larger in the lower leg muscles. The FF_{MR1} values ranged from 6.3% ± 4.4% (TP) to 30.8% ± 22.4% (GM), while $FF_{MR\ 2}$ values were very similar and ranged from 6.2% ± 4.2% (TP) to 30.5% ± 22.2% (GM). The corresponding ICC values were high except for the adductor magnus (0.61), and a most values were superior to 0.90. They ranged from 0.61 to 0.99 for the thigh muscles and from 0.95 to 1.00 for the lower leg muscles. All muscles, except GL, have a lower ICC value for the measurement repeatability segmentation than for the automatic segmentation versus the manual segmentation.

3.3 | Interexpert repeatability

The average FF values and the corresponding ICC and SEM values for the interexpert repeatability assessment are found in Table 4. For the thigh muscles, FF_M values ranged from 7.1% ± 4.0% (VM) to 31.4% ± 34.9% (BFS), and FF_{A1} values ranged from 7.1% ± 3.9% (VM) to 31.7% ± 35.3% (BFS), which is quite similar to the FF_{A2} values that ranged from 7.1% ± 3.9% (VM) to 31.4% ± 36.0% (BFS). In the lower leg, FF_M values ranged from 8.5% ± 7.1% (GL) to 34.2% ± 34.9% (GM), and FF_{A1} values ranged from 13.4% ± 9.1% (GL) to 34.8% ± 34.7% (GM), with FF_{A2} values quite similar ranging from 10.6% ± 8.0% (GL) to 34.5% ± 35.2% (GM).

For both experts, the FF values calculated using the semi-automatic method compared with those obtained using manual segmentation of the main expert of the study provided ICC values higher than 0.95 except for the adductor magnus (0.90 and 0.91 for FF_{A1} and FF_{A2}, respectively) and the GL (0.83 and 0.84 for FF_{A1} and FF_{A2}, respectively). For the DSC scoring, comparing the initial segmentations of both experts resulted in DSC scores ranging from 0.69 to 0.96 (Supporting Information Table S1). In addition, comparing the propagations resulting from the manual segmentations of expert 2 with the full manual segmentation of expert 1 gave DSC scores ranging from 0.75 to 0.95.

TABLE 3 Mean FF values, ICC, and SEM measured at 2 different times for each muscle of a group of 4 subjects

Muscle	FF_{MR1} (%)	FF_{MR2} (%)	ICC	SEM (%)
AL	5.6 ± 3.2	5.8 ± 3.4	0.97	0.6
AM	5.8 ± 3.2	4.8 ± 2.6	0.61	1.8
BFL	7.4 ± 4.5	6.6 ± 3.8	0.93	1.0
BFS	11.3 ± 8.4	10.4 ± 7.8	0.99	0.6
G	6.7 ± 3.3	7.2 ± 4.4	0.93	1.0
RF	5.4 ± 2.7	6.2 ± 4.3	0.88	1.2
S	10.7 ± 6.0	11.3 ± 8.1	0.93	1.9
SM	8.7 ± 5.6	9.0 ± 5.5	0.92	1.6
ST	6.8 ± 3.7	6.8 ± 4.1	0.94	0.9
VI	7.4 ± 5.5	7.4 ± 5.8	0.99	0.5
VL	6.0 ± 3.2	6.5 ± 3.9	0.94	0.9
VM	5.3 ± 2.9	5.8 ± 3.9	0.89	1.2
EDL-EHL	9.6 ± 5.4	9.5 ± 4.8	0.97	0.9
FB-FL	22.5 ± 19.5	21.8 ± 19.7	1.00	0.7
GL	12.6 ± 16.5	11.9 ± 14.8	0.99	1.3
GM	30.8 ± 22.4	30.5 ± 22.2	1.00	0.3
SOL	22.9 ± 24.2	22.5 ± 24.3	1.00	0.8
TA	7.2 ± 4.6	7.2 ± 4.1	0.95	0.9
TP	6.3 ± 4.4	6.2 ± 4.2	0.97	0.8

Note: Segmentation was performed manually twice (FF_{MR1} and FF_{MR2}). The ICC and SEM values were computed for each muscle on the basis of a comparative analysis between both methods.

4 | DISCUSSION

A semi-automatic segmentation tool of the whole set of individual thigh and lower leg muscles was designed and assessed in patients with neuromuscular disorders, both cross-sectionally and longitudinally.

The present method allowed the segmentation of 19 individual muscles in MR images of patients based on the manual delineation of these muscles on 9 appropriately chosen slices. The corresponding masks were used as seed regions, which were transversally and longitudinally propagated to the remaining 91 slices for both the baseline and follow-up data sets. The corresponding DSC values ranged from 0.83 to 0.96 for the transversal propagation and from 0.76 to 0.94 for the longitudinal propagation. The few studies in which the issue of image segmentation from patients with severe fat infiltration has been addressed only aimed at distinguishing muscle from fat compartments.[21-24] None of them addressed individual muscles segmentation for which fat infiltration and/or muscle atrophy are definitive confounding factors. Individual muscle segmentation has been performed in healthy subjects and patients with chronic obstructive pulmonary disease, 2 groups of subjects in whom fat infiltration is moderate.[14,15] In MR images from healthy subjects, Baudin et al reported a global mean DSC of 0.86 ± 0.07, whereas in images from chronic obstructive pulmonary disease patients, Andrews et al reported DSC values ranging from 0.70 ± 0.16 to 0.93 ± 0.06. It should be noted that the algorithm of Andrews et al performed significantly worse when muscle boundaries were less visible, which is often the case for severely fat-infiltrated muscles in neuromuscular disorders and performed

TABLE 4 Mean FF values, ICC, and SEM measured for 4 baseline data sets of the thighs and legs

		Expert 1			Expert 2		
Muscle	**FF_M (%)**	**FF_{A1} (%)**	**ICC**	**SEM (%)**	**FF_{A2} (%)**	**ICC**	**SEM (%)**
AL	7.5 ± 4.4	7.4 ± 4.2	0.99	0.4	7.3 ± 4.3	1.00	0.2
AM	11.2 ± 7.7	14.2 ± 8.0	0.90	2.5	13.8 ± 8.0	0.91	2.3
BFL	20.3 ± 20.3	20.4 ± 20.5	1.00	0.2	20.4 ± 20.3	1.00	0.1
BFS	31.4 ± 34.9	31.7 ± 35.3	1.00	0.4	31.4 ± 36.0	1.00	2.0
G	10.8 ± 6.5	10.6 ± 6.3	1.00	0.3	10.8 ± 6.5	1.00	0.1
RF	7.9 ± 4.9	8.4 ± 4.9	0.98	0.7	8.6 ± 4.9	0.95	1.1
S	19.0 ± 14.8	19.0 ± 14.7	1.00	0.1	19.1 ± 14.1	1.00	0.9
SM	21.6 ± 22.1	22.1 ± 22.6	1.00	0.4	20.9 ± 20.8	1.00	1.1
ST	26.3 ± 32.6	26.2 ± 32.6	1.00	0.3	25.9 ± 33.0	1.00	0.7
VI	9.9 ± 6.9	10.9 ± 7.2	0.98	1.1	10.9 ± 6.9	0.96	1.3
VL	9.9 ± 5.0	10.0 ± 4.9	1.00	0.2	10.1 ± 5.3	0.99	0.4
VM	7.1 ± 4.0	7.1 ± 3.9	1.00	0.1	7.1 ± 3.9	1.00	0.1
EDL-EHL	14.2 ± 13.4	14.6 ± 13.6	1.00	0.2	14.3 ± 13.2	1.00	0.8
FB-FL	8.5 ± 7.1	13.4 ± 9.1	0.83	3.3	10.6 ± 8.0	0.84	3.0
GL	34.2 ± 34.9	34.8 ± 34.7	1.00	0.7	34.5 ± 35.2	1.00	1.4
GM	28.7 ± 33.3	28.6 ± 33.1	1.00	0.3	26.7 ± 34.2	1.00	1.6
SOL	20.0 ± 26.9	20.1 ± 26.6	1.00	0.3	20.2 ± 26.2	1.00	0.6
TA	20.4 ± 28.8	20.7 ± 29.1	1.00	0.2	20.9 ± 29.4	1.00	0.4
TP	16.9 ± 12.1	16.9 ± 12.4	1.00	0.2	16.6 ± 12.4	0.99	1.0

Note: Mean FFs were quantified on masks resulting from the manual (FF_M) segmentation and the automatic (FF_A) segmentations initiated by 2 different experts for each muscle (FF_{A1} and FF_{A2}, respectively). The ICC and SEM values were computed for each muscle from a comparative analysis between manual and automatic methods.

differently depending on the MRI modality used. In the present study, the whole set of MR modalities tested did provide DSC values larger than 0.80 and the water contrast was defined as the optimal modality.

One of our aims was to assess the potential of our segmentation tool for follow-up studies in patients with neuromuscular disorders. In that respect, in addition to the conventional DSC metric, we assessed one of the most commonly reported outcome measures in this field: the FF. Such an approach has already been reported by Gadermayr et al,[23] who used the FF error computed in T_1-weighted images as an additional metric to assess the performance of their automatic segmentation method of muscle tissue as a whole. To go further, we assessed the performance of our segmentation framework from FF of all individual muscles measured from proton density fat fraction images, which are quantitative images in contrast to T_1-weighted images that only provide information related to contrast. Our results demonstrated a very high reliability for both the transversal (ICC > 0.98) and longitudinal propagation (ICC > 0.94) for each individual muscle. In addition, we found higher ICC values for our semi-automatic method as compared with the measurement repeatability. This strongly supports that, regarding fat infiltration quantification, the variation in FF caused by automatic segmentation is comparable to closeness of the agreement between 2 successive measurements on the same day. Furthermore, high ICC values and DSC values were also reported for the transversal propagation of expert 2 compared with the ground-truth manual segmentation of expert 1, which demonstrates a good interexpert repeatability. Together, this demonstrates that our proposed algorithm can result in reliable and repeatable FFs, which is essential for follow-up measurements.

The increase in FF we observed from baseline to follow-up is related to the natural progression of the disease, as also previously reported for a large number of neuromuscular diseases such as Duchenne muscular dystrophy[34] and fascioscapulohumeral dystrophy.[12]

Regarding the transversal propagation, an initial manual segmentation is mandatory and can be considered a limiting factor considering the well-known operator dependency.[7] This process is even more difficult in the case of severely infiltrated muscles, for which the delimitation of muscle boundaries might be difficult. In that respect, inappropriate 2D manual segmentations might be propagated, thereby compromising the 3D automatic segmentation. This issue is especially relevant for muscles showing large anatomical variations along the leg. Those muscles for which the transversal propagation was suboptimal also showed a suboptimal

longitudinal propagation, thereby illustrating the interdependence of both methods. The semi-automatic nature of our method and the need for manual segmentation input offer a large flexibility regarding the number of muscles to be segmented. One might be interested in whole-muscle compartments, while others might intend to segment a variable number of individual muscles. Regardless of the type of inputs, corresponding masks can be propagated to the whole 3D region using the present method, not only at baseline, but also for an unlimited number of repeated acquisitions.

Regarding the longitudinal propagation, segmentation results were very high for almost all of the subjects. We observed that suboptimal results were obtained when muscles shapes were changing between the baseline and the follow-up data sets. Considering that muscles are made of soft tissue, one can expect large shape changes due to various constraints imposed by leg and/or coil positioning. As illustrated in Figure 3, our longitudinal propagation process performed well despite these potential anatomical variations, whether they were global crushing pressures (P04) or localized pressure points (P02). As illustrated by the GL muscle in subject P10, a localized pressure point can be such that the muscle shape can change dramatically, thereby compromising the automatic segmentation process. This issue can also be relevant for follow-up studies based on manual segmentation only. Overall, distortions regarding muscle shapes for repeated acquisitions represent biases that would have to be accounted for in future studies.

The high DSC values obtained in the present study as compared with previously reported methods[14,15] and the very high reliability of FF values strongly support that our segmentation framework can be considered reliable and accurate for the segmentation of individual muscles in patients with neuromuscular disorders and so for both cross-sectional as longitudinal studies.

5 | CONCLUSIONS

We present a reliable semi-automatic method for segmentation of individual muscles in MR images from patients with neuromuscular disorders. It relies on propagation methods based on nonlinear registration and is effective for cross-sectional and longitudinal studies. As far as we know, this is the very first segmentation framework that is dedicated to individual muscle segmentation in neuromuscular diseases ever reported for follow-up studies, which are of utmost importance in this clinical field. Both the DSC and FF values demonstrated the high performance of our method, which allowed a substantial reduction of the number of slices to be manually segmented. With no additional manual input, follow-up scans can be processed automatically with a very high reliability. This minimal manual segmentation stage, which allows the choice of regions of interest, should encourage clinicians to use such a method in a clinical environment for a quantitative assessment of muscular changes occurring in neuromuscular disorders.

ACKNOWLEDGMENTS

The authors thank all of the subjects who participated in the present study. This study was supported by the EU Seventh Framework Program on DM1 (OPTIMISTIC) (grant/award No. 305697).

ORCID

Augustin C. Ogier https://orcid.org/0000-0001-9178-9964

REFERENCES

1. Janssen BH, Voet NBM, Nabuurs CI, et al. Distinct disease phases in muscles of facioscapulohumeral dystrophy patients identified by MR detected fat infiltration. Kassubek J, ed. *PLoS ONE*. 2014;9:e85416.
2. Wokke BH, van den Bergen JC, Versluis MJ, et al. Quantitative MRI and strength measurements in the assessment of muscle quality in Duchenne muscular dystrophy. *Neuromuscular Disord*. 2014;24:409–416.
3. Morrow JM, Sinclair CDJ, Fischmann A, et al. MRI biomarker assessment of neuromuscular disease progression: a prospective observational cohort study. *Lancet Neurol*. 2016;15:65–77.
4. Arpan I, Willcocks RJ, Forbes SC, et al. Examination of effects of corticosteroids on skeletal muscles of boys with DMD using MRI and MRS. *Neurology*. 2014;83:974–980.
5. Janssen B, Voet N, Geurts A, van Engelen B, Heerschap A. Quantitative MRI reveals decelerated fatty infiltration in muscles of active FSHD patients. *Neurology*. 2016;86:1700–1707.
6. Mercuri E, Pichiecchio A, Allsop J, Messina S, Pane M, Muntoni F. Muscle MRI in inherited neuromuscular disorders: past, present, and future. *J Magn Reson Imaging*. 2007;25:433–440.
7. Barnouin Y, Butler-Browne G, Voit T, et al. Manual segmentation of individual muscles of the quadriceps femoris using MRI: a reappraisal. *J Magn Reson Imaging*. 2014;40:239–247.
8. Carlier PG, Marty B, Scheidegger O, et al. Skeletal muscle quantitative nuclear magnetic resonance imaging and spectroscopy as an outcome measure for clinical trials. *J Neuromusc Dis*. 2016;3:1–28.
9. Hooijmans MT, Niks EH, Burakiewicz J, et al. Non-uniform muscle fat replacement along the proximodistal axis in Duchenne muscular dystrophy. *Neuromusc Disord*. 2017;27:458–464.
10. Kornblum C, Lutterbey G, Bogdanow M, et al. Distinct neuromuscular phenotypes in myotonic dystrophy types 1 and 2: a whole body highfield MRI study. *J Neurol*. 2006;253:753–761.
11. Lareau-Trudel E, Le Troter A, Ghattas B, et al. Muscle quantitative MR imaging and clustering analysis in patients with facioscapulohumeral muscular dystrophy type 1. Fan Y, ed. *PLoS ONE*. 2015;10:e0132717.
12. Fatehi F, Salort-Campana E, Le Troter A, et al. Long-term follow-up of MRI changes in thigh muscles of patients with

Facioscapulohumeral dystrophy: a quantitative study. Kyba M, ed. *PLoS ONE*. 2017;12:e0183825.
13. Essafi S, Langs G, Deux J-F, Rahmouni A, Bassez G,Paragios N. Wavelet-driven knowledge-based MRI calf muscle segmentation. In: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, 2009. pp 225–228.
14. Baudin PY, Azzabou N, Carlier PG, Paragios N. Prior knowledge, random walks and human skeletal muscle segmentation. *Med Image Comput Comput Assist Interv*. 2012;15(Pt 1):569–576.
15. Andrews S, Hamarneh G. The generalized log-ratio transformation: learning shape and adjacency priors for simultaneous thigh muscle segmentation. *IEEE Trans Med Imaging*. 2015;34:1773–1787.
16. Ahmad E, Yap MH, Degens H, McPhee JS. Atlas-registration based image segmentation of MRI human thigh muscles in 3D space. In: Proceedings of Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment, San Diego, CA, 2014: 9037.
17. Karlsson A, Rosander J, Romu T, et al. Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI: automated muscle tissue quantification. *J Magn Reson Imaging*. 2015;41:1558–1569.
18. Le Troter A, Fouré A, Guye M, et al. Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches. *MAGMA*. 2016;29:245–257.
19. Positano V, Christiansen T, Santarelli MF, Ringgaard S, Landini L, Gastaldelli A. Accurate segmentation of subcutaneous and intermuscular adipose tissue from MR images of the thigh. *J Magn Reson Imaging*. 2009;29:677–684.
20. Orgiu S, Lafortuna CL, Rastelli F, Cadioli M, Falini A, Rizzo G. Automatic muscle and fat segmentation in the thigh from T1-weighted MRI: automatic T_1-W MRI segmentation of the thigh. *J Magn Reson Imaging*. 2016;43:601–610.
21. Kovacs W, Liu C-Y, Summers RM, Yao J. Identification of muscle and subcutaneous and intermuscular adipose tissue on thigh MRI of muscular dystrophy. In: Proceedings of IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 2016. pp 176–179.
22. Yao J, Kovacs W, Hsieh N, Liu C-Y, Summers RM. Holistic segmentation of intermuscular adipose tissues on thigh MRI. In: Proceedings of the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, Canada, 2017. pp 737–745.
23. Gadermayr M, Disch C, Müller M, Merhof D, Gess B. A comprehensive study on automated muscle segmentation for assessing fat infiltration in neuromuscular diseases. *Magn Reson Imaging*. 2017;48:20–26.
24. Mesbah S, Shalaby A, Stills S, et al. A novel automatic segmentation method to quantify the effects of spinal cord injury on human thigh muscles and adipose tissue. In: Proceedings of the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, Canada, 2017. Vol. 10434.
25. Ogier A, Sdika M, Foure A, Le Troter A, Bendahan D. Individual muscle segmentation in MR images: a 3D propagation through 2D non-linear registration approaches. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 2017. pp 317–320.
26. Okkersen K, Jimenez-Moreno C, Wenninger S, et al. Cognitive behavioural therapy with optional graded exercise therapy in patients with severe fatigue with myotonic dystrophy type 1: a multicentre, single-blind, randomised trial. *Lancet Neurol*. 2018;17:671–680.
27. Mathieu J, Boivin H, Meunier D, Gaudreault M, Bégin P. Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy. *Neurology*. 2001;56:336–340.
28. Jenkinson M, Beckmann CF, Behrens T, Woolrich MW, Smith SM. FSL. *Neuroimage*. 2012;62:782–790.
29. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29:1310–1320.
30. Tustison NJ, Avants BB. Explicit B-spline regularization in diffeomorphic image registration. *Front Neuroinform*. 2013;7.
31. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297–302.
32. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
33. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003;22:85–93.
34. Hollingsworth KG, Garrood P, Eagle M, Bushby K, Straub V. Magnetic resonance imaging in duchenne muscular dystrophy: longitudinal assessment of natural history over 18 months: short reports. *Muscle Nerve*. 2013;48:586–588.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

TABLE S1 Dice similarity coefficients measured for 4 baseline data sets of the thighs and legs to assess the interobserver variability. Note: The DSCs were obtained by comparing the ground-truth segmentation (Manual) with the segmentations of 2 different experts used to initiate the automatic method (IniAuto) and the results of the automatic method (Auto)

How to cite this article: Ogier AC, Heskamp L, Michel CP, et al. A novel segmentation framework dedicated to the follow-up of fat infiltration in individual muscles of patients with neuromuscular disorders. *Magn Reson Med*. 2019;00:1–12. https://doi.org/10.1002/mrm.28030