# Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches

Arnaud Le Troter, Alexandre Fouré, Maxime Guye, Sylviane Confort-Gouny, Jean-Pierre Mattei, Julien Gondin, Emmanuelle Salort-Campana, David Bendahan

# Volume measurements of individual muscles in human *quadriceps femoris* using atlas-based segmentation approaches

**Arnaud Le Troter[1,2] · Alexandre Fouré[1,2] · Maxime Guye[1,2] · Sylviane Confort-Gouny[1,2] · Jean-Pierre Mattei[1,3] · Julien Gondin[1,2] · Emmanuelle Salort-Campana[4] · David Bendahan[1,2]**

Received: 7 July 2015 / Revised: 11 February 2016 / Accepted: 12 February 2016

**Abstract**

*Objectives* Atlas-based segmentation is a powerful method for automatic structural segmentation of several sub-structures in many organs. However, such an approach has been very scarcely used in the context of muscle segmentation, and so far no study has assessed such a method for the automatic delineation of individual muscles of the *quadriceps femoris* (QF). In the present study, we have evaluated a fully automated multi-atlas method and a semi-automated single-atlas method for the segmentation and volume quantification of the four muscles of the QF and for the QF as a whole.

*Subjects and methods* The study was conducted in 32 young healthy males, using high-resolution magnetic resonance images (MRI) of the thigh. The multi-atlas-based segmentation method was conducted in 25 subjects. Different non-linear registration approaches based on free-form deformable (FFD) and symmetric diffeomorphic normalization algorithms (SyN) were assessed. Optimal parameters of two fusion methods, i.e., STAPLE and STEPS, were determined on the basis of the highest Dice similarity index (DSI) considering manual segmentation (MSeg) as the ground truth. Validation and reproducibility of this pipeline were determined using another MRI dataset recorded in seven healthy male subjects on the basis of additional metrics such as the muscle volume similarity values, intraclass coefficient, and coefficient of variation. Both non-linear registration methods (FFD and SyN) were also evaluated as part of a single-atlas strategy in order to assess longitudinal muscle volume measurements. The multi- and the single-atlas approaches were compared for the segmentation and the volume quantification of the four muscles of the QF and for the QF as a whole.

*Results* Considering each muscle of the QF, the DSI of the multi-atlas-based approach was high 0.87 ± 0.11 and the best results were obtained with the combination of two deformation fields resulting from the SyN registration method and the STEPS fusion algorithm. The optimal variables for FFD and SyN registration methods were four templates and a kernel standard deviation ranging between 5 and 8. The segmentation process using a single-atlas-based method was more robust with DSI values higher than 0.9. From the vantage of muscle volume measurements, the multi-atlas-based strategy provided acceptable results regarding the QF muscle as a whole but highly variable results regarding individual muscle. On the contrary, the performance of the single-atlas-based pipeline for individual muscles was highly comparable to the MSeg, thereby indicating that this method would be adequate for longitudinal tracking of muscle volume changes in healthy subjects.

*Conclusion* In the present study, we demonstrated that both multi-atlas and single-atlas approaches were relevant for the segmentation of individual muscles of the QF in

A. Le Troter and A. Fouré contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10334-016-0535-6) contains supplementary material, which is available to authorized users.

✉ Arnaud Le Troter
arnaud.le-troter@univ-amu.fr

[1] Aix Marseille Université, CNRS, CRMBM UMR 7339, 13385 Marseille, France

[2] APHM, CHU Timone, Pôle imagerie médicale, CEMEREM, 13005 Marseille, France

[3] APHM, CHU Sainte-Marguerite, Département de Rhumatologie, 13009 Marseille, France

[4] APHM, CHU Timone, Centre de Référence des Maladies Neuromusculaires et de la SLA, 13005 Marseille, France

Published online: 16 March 2016

healthy subjects. Considering muscle volume measurements, the single-atlas method provided promising perspectives regarding longitudinal quantification of individual muscle volumes.

**Keywords** MRI · Multi-atlas-based segmentation · *Quadriceps femoris* muscle · Non-linear registration · Fusion · Individual muscle volume measurements

## Introduction

Segmentation of skeletal muscle magnetic resonance (MR) images is of high interest given that it can provide accurate quantitative information regarding muscle volume, fat infiltration, and the corresponding changes in a variety of physiological (e.g., exercise, immobilization, aging) and pathological conditions (e.g., neuromuscular diseases). If such an approach can be robust, reproducible, and automated, it can be promising for the processing of large amounts of data such as those obtained in trials for which accurate longitudinal measurements have to be performed. However, considering the high variability of muscles shapes and the relative positions within the image, segmentation of muscle MR images has been commonly recognized as challenging, generally speaking, and even more so if one aims at quantifying the volume of each muscle. More particularly for the thigh muscles, any subtle changes regarding the angular position of the leg, the relative position of the coil on the thigh, and the stress of the coil on the thigh can generate, among other things, additional uncontrolled factors of variability. In order to circumvent these difficulties, manual labelling of anatomical structures has been used in several studies, and the time consuming and expert-dependent nature of the task have been largely recognized [1]. Recently, several semi-automated or automated segmentation methods have been proposed for skeletal muscle MR images parcellation.

Baudin et al. [2, 3] associated a statistical shape atlas and an random walks graph-based algorithm in order to automatically segment individual muscles, and reported high Dice similarity indices (DSI $= 0.86 \pm 0.07$) for 13 thigh muscles. Based on a combined analysis of water and fat volumetric MRI of the mid-thigh region, Makrogiannis et al. [4] validated an unsupervised muscle and fat quantification algorithm. In order to take into account the large intersubject variability, Prescott et al. [5] presented a semi-automated method of segmenting each muscle of the *quadriceps femoris* (QF) based on the manual selection of specific anatomical landmarks and the utilisation of a selected template from a templates database. They assessed the performance of this approach as compared to a manual method and reported the following values: *rectus femoris* (RF), DSI $= 0.78 \pm 0.12$; *vastus intermedius* (VI), DSI $= 0.79 \pm 0.10$; *vastus lateralis* (VL), DSI $= 0.82 \pm 0.08$; and *vastus medialis* (VM), DSI $= 0.69 \pm 0.16$. Using a combination of a manual labelling of regions of interest, atlas construction, and registration, Ahmad et al. [6] proposed a semi-ASeg pipeline of the QF that proved to be reliable. However, the authors did not provide information related to individual muscles of the QF. Overall, considering the large intersubject variability regarding individual muscle morphology and size, one could hypothesize that a multi-atlas-based segmentation would more accurately take into account this anatomical variability and provide more robust results.

Accordingly, a registration process using multiple image templates on the one hand and a subsequent label fusion on the other hand proved to be reliable for the in vivo parcellation of a mouse brain [7]. A similar performance has been reported for human brain segmentation using label fusion and multi-atlas-based registration [8–10]. It is noteworthy that a multi-atlas-based approach has also been reported for the prostate [11] and pectoral muscles [12]. More recently, a few studies have investigated the potential of multi-atlas-based approaches in order to automatically quantify large muscle volumes from whole-body MR images [13, 14]. Overall, although of interest, multi-atlas-based methods have been mainly used for large muscle volumes measurements and the potential of such an approach for the quantification of individual muscle volumes and its corresponding accuracy for longitudinal measurements have never been assessed.

From the vantage of robustness, most of the studies reported so far have used DSI values higher than 0.7 as a sign of an acceptable performance [15] and muscle volume measurements have been scarcely reported. Additional metrics such as the false positive (FPVF) and the false negative volume fraction (FNVF) have been advocated as reliable indices of the accuracy of segmentation and volume measurements [16].

The purpose of the present study was therefore to assess and compare the performance of fully automated, multi-atlas-based and semi-automated single-atlas-based algorithms for the segmentation and volume quantification of each muscle of the QF in healthy subjects, as well as to characterize the corresponding accuracy for longitudinal measurements. We also aimed at comparing the performance of different non-linear registration and fusion processes as part of the atlas-based pipeline.

## Materials and methods

### Subjects

Two groups of subjects were investigated after they were fully informed about the nature and aim of the study and gave their informed written consent. The first group of subjects was composed of 25 healthy men (22 ± 1 years, height 178 ± 6 cm, weight 68 ± 7 kg) while seven healthy men (32 ± 7 years, height 169 ± 9 cm, weight 64 ± 11 kg) were part of the second group. A single MRI acquisition was performed in the first group of subjects while two MRI sessions was repeated twice, 2 days apart, in the second group of subjects. The study was approved by the local human research ethics committee and was conducted in conformity with the Declaration of Helsinki.

### In vivo MRI data acquisition

Subjects were positioned supine with the right leg centered in a 1.5 T superconducting magnet (MAGNETOM Avanto, Siemens AG, Healthcare Sector, Erlangen, Germany). A flexible surface 6-channel body coil (Siemens AG, Healthcare Sector, Erlangen, Germany) was placed around the right thigh. $T_1$-weighted ($T_{1w}$) high-resolution images (20 slices, field of view = 220 mm × 220 mm; matrix = 576 × 576; time repetition = 549 ms; echo time = 13 ms; number of repetitions = 1; slice thickness = 6 mm; gap between slices = 6 mm, acquisition time = 5 min 18 s) were recorded using a gradient-echo sequence. The most distal slice was always acquired at approximately 20 mm (i.e. 5 % of the thigh length measured for each subject) upper the proximal border of the patella.

As previously described by Barnouin et al. [1], borders of the anatomic cross sectional areas of the four QF muscles, i.e. the VL, VM, VI, and RF, were manually drawn in each slice by an experienced researcher (A.F., with 7 years of experience in evaluation of muscle anatomy and geometry). The corresponding result was referred to as manual segmentation (MSeg) in the following sections and was considered as the ground truth. All the regions of interest were delineated using FSL View software, the 3D viewer included in the FSL toolbox [17].

### Metrics for the segmentation process evaluation

Four metrics (ranging from 0 to 1) were used in order to evaluate the segmentation algorithms on the basis of a comparative analysis between the results of an ASeg ($X$) and those from an MSeg ($Y$), performed by an expert (considered the ground truth):

- The Dice similarity index $\mathrm{DSI}(X,Y) = \frac{2\times|X\cap Y|}{|X|+|Y|}$ is a quantitative index of the overlap between $X$ and $Y$
- The false negative volume fraction FNVF ($X$, $Y$) refers to the volume fraction corresponding to the missed parts of the ground truth $Y$ into $X$
- The false positive volume fraction FPVF ($X$, $Y$) refers to the volume fraction corresponding to the segmented parts not overlapping the ground truth into $X$.
- The muscle volume similarity fraction $\mathrm{MVSF}(X,Y) = \frac{2\times|v(X)-v(Y)|}{v(X)+v(Y)}$ refers to the similarity between the manually $v(Y)$ and the automatically $v(X)$ quantified volumes.

### Atlas construction

As illustrated in Fig. 1, a single atlas was built from of a $T_{1w}$ image (Fig. 1a) and the corresponding segmentations of specific regions of interest, i.e. intermuscular fat (IMAT), subcutaneous fat (SAT), muscle (MUSCLE), and bone, using an entirely automated segmentation algorithm (ASeg). ASeg was based on an initial pixel-based intensity analysis of the

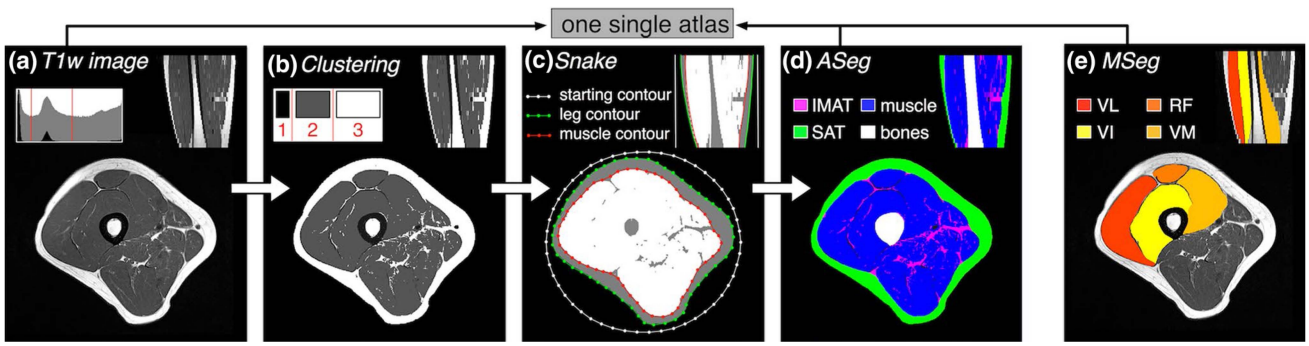


**Fig. 1** Main steps of a single-atlas construction including an automatic segmentation (ASeg) of intramuscular adipose tissue (IMAT), muscle, subcutaneous adipose tissue (SAT), bone, and a manual segmentation (MSeg) of *vastus lateralis* (*VL*), *rectus femoris* (*RF*), *vastus medialis* (*VM*), and *vastus intermedius* (*VI*)

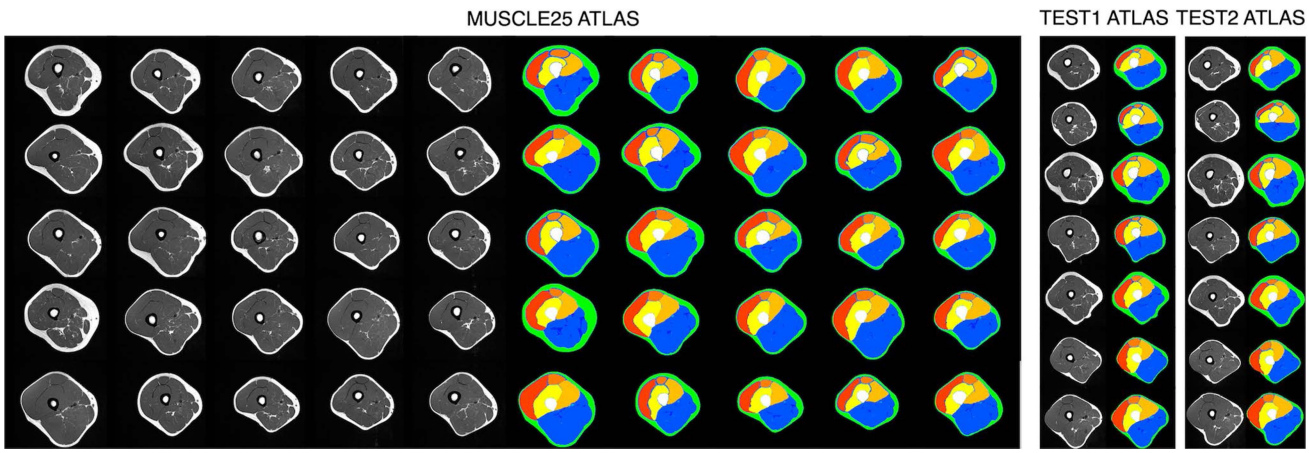


**Fig. 2** $T_1$-weighted images ($T_{1w}$) recorded in both groups of subjects and the corresponding atlases MUSCLE25, TEST1, and TEST2 [colour legend: vastus lateralis (*red*), rectus femoris (*dark orange*), vastus medialis (*light orange*) and vastus intermedius (*yellow*), bone (*white*), subcutaneous adipose tissue (*green*), intermuscular adipose tissue (*dark blue*), posterior thigh (*blue*)]

$T_{1w}$ image (Fig. 1b) followed by a spatial analysis allowing us to identify each region of interest using a colour label (more details are provided in "Automated segmentation: ASeg"). Overall, the initial $T_{1w}$ image together with the results of both segmentation processes ASeg (Fig. 1d) and MSeg (Fig. 1e) include eight labels and refer to a single atlas.

*Automated segmentation: ASeg*

ASeg was intended to obtain a fast and fully automated segmentation method providing quantitative information related to muscle, bone, and fat fractions (intermuscular and subcutaneous). It has been developed in C++ using the OpenCV library previously published by Positano et al. [18]. Briefly, the first step of the ASeg method corresponds to a *k*-means clustering applied on the intensity values of each voxel in order to segment the original image into three clusters (Fig. 1b, top left corner) corresponding to background (cluster 1), muscle tissue (cluster 2), and adipose tissue (cluster 3). In the second step, we used a morphological closing with a small structuring element in order to eliminate noise from the vessels in the SAT region, which keeps the topology of the leg and muscle regions. Then, clusters 2 and 3 (muscle and adipose tissue) were merged to obtain a new shape corresponding to the leg. At this stage, an active polygonal contour was performed in cluster 1 (the white discrete circle on Fig. 1c) in order to define the boundaries between the leg and the muscle (the red discrete contour on Fig. 1c). While Positano et al. [18] used a gradient vector flow snake algorithm previously described by Xu et al. [19], we used a more classical polygonal active contour algorithm described by the following energy with the aim of minimizing $\varepsilon$:

$$\varepsilon = \oint_{\text{contour}C(s)} (\alpha(s)E_{\text{cont}} + \beta(s)E_{\text{curv}} + \gamma(s)E_{\text{image}})\mathrm{d}s$$

where $E_{\text{cont}} = \|p_i - p_{i-1}\|^2$ and $E_{\text{curv}} = \|p_{i-1} - 2p_i + p_{i+1}\|^2$ represent the internal energy related, respectively, to the continuity and the curvature of the shape contour (defined as $C_{0 \leqq i < N} = \{p_i\}$ where $N = 50$), and $E_{\text{image}}$ represents the external energy on the contour of the binary target image. The parameters ($\alpha$, $\beta$, $\gamma$) control the contour tension, the rigidity of the curve, and the data attachment. In the present study, we chose the three weighting parameters (0.3, 0.4, 0.3).

The MRI datasets recorded in the two groups of subjects were used in order to build three different atlases. The MUSCLE25 atlas was composed of the 25 subjects of group 1. The repeated measurements performed in the second group of subjects allowed us to build two additional atlases referred as TEST1 and TEST2. These three atlases were used in order to assess both segmentation methods proposed in this study (Fig. 2). For the fully automated, multi-atlas approach, TEST1 and TEST2 have been defined as the images to be segmented (targets) and MUSCLE25 as the source atlas. For the semi-automated single-atlas approach, the MUSCLE25 atlas has not been used, and TEST1 and TEST2 have been used as source and target atlases and vice versa.

## Fully automated multi-atlas based segmentation

The global scheme of the fully automated segmentation pipeline is described in Fig. 3. It includes a first stage dedicated to the non-linear registration algorithm defined as the $T$ function (more details are provided in "Affine and non-linear registration methods") and a second stage dedicated to a multi-atlas fusion approach defined as the $F$ function (more details are provided in "Multi-atlas label fusion methods"). The performance of this pipeline

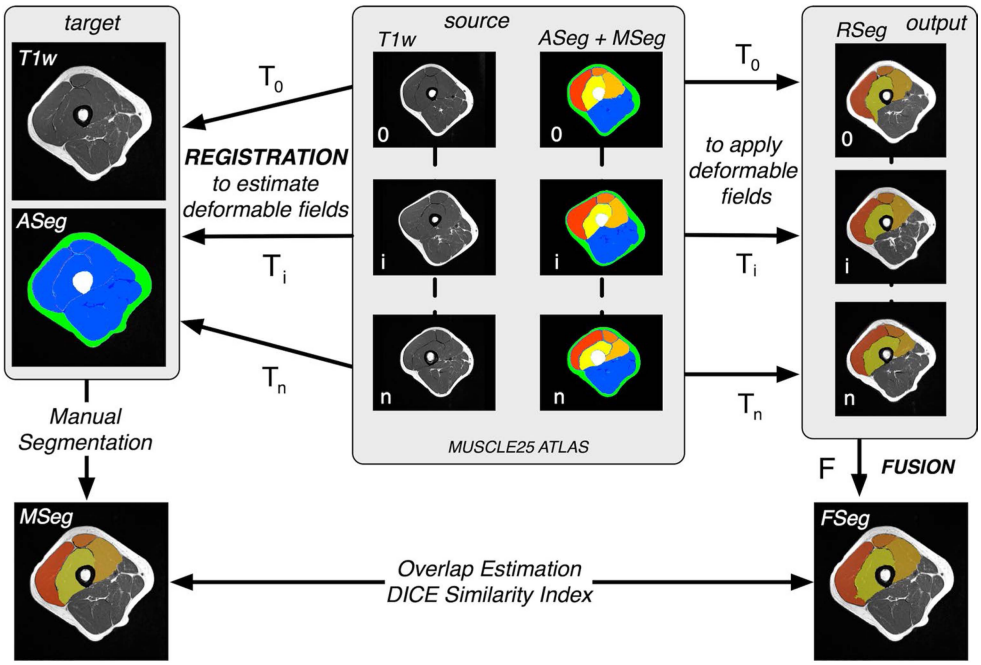


**Fig. 3** Automated segmentation pipeline using the multi-atlas-based registration and fusion algorithms. *T* refers to a non-linear registration function, *MSeg* to the manual segmentation of $T_{1w}$ image, *ASeg* to the automated segmentation of $T_{1w}$, *RSeg* to the automated segmentation without the fusion process, *FSeg* to the automated segmentation resulting from the fusion function (*F*)

was assessed using MUSCLE25 as the source atlas and TEST1 and TEST2 as targets, i.e. set of images to be segmented.

### Semi-automated single-atlas based segmentation

The global scheme of the semi-automated segmentation pipeline is described in Fig. 4. It includes a single stage dedicated to the non-linear registration (defined as the *T* function). The performance of this pipeline was assessed using TEST1 and TEST2 as source and targets and vice versa. More specifically, each image set of a given subject in TEST1 was registered using the image set of the same subject in TEST2 and vice versa. This method can only be used for multiple MRI recordings in the same subject. The result of the MSeg performed for the first MRI is used as the single-atlas source for the segmentation of other MRI recordings considered as the targets.

### Affine and non-linear registration methods

In order to take into account the differences between the target image (image to be segmented) and the atlases, we assessed two different approaches, i.e. the free-form deformable (FFD) parametric method originally proposed by Rueckert et al. [20] and implemented by Modat et al. (available in the library *NiftiReg* [21]), and the non-parametric symmetric diffeomorphic normalization (SyN) method, implemented in the *ANTs* library [22].

In order to facilitate an understanding of the following parts, we have detailed below a few abbreviations:

- $\mathrm{RSeg}^{\mathrm{reg}}_{I,J} = T(I,J)_{\mathrm{reg}}$ refers to the resulting image of the registration process between the source image $I$ and the target image $J$.
- $T(I, J)$ refers to the transformation describing the deformation field between $I$ and $J$.

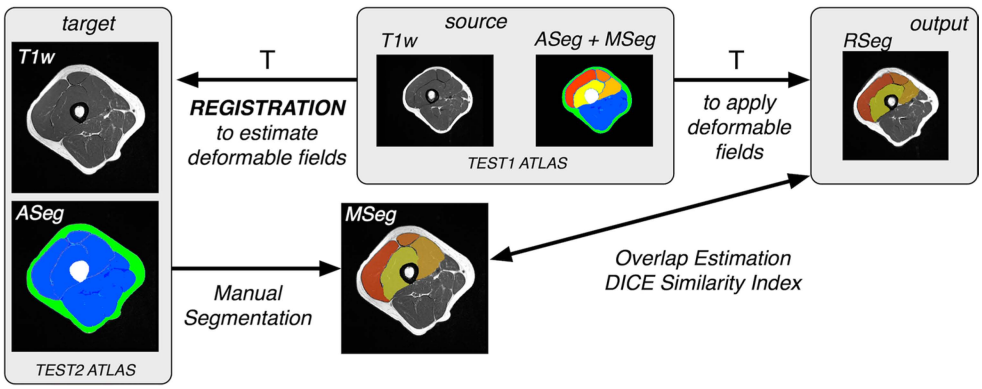


**Fig. 4** Semi-automated segmentation pipeline using the single-atlas based registration. *T* refers to a non-linear registration function, *MSeg* to the manual segmentation of $T_{1w}$ image, *ASeg* to the automated segmentation of $T_{1w}$, *RSeg* to the automated segmentation resulting from the registration function (*T*)

- reg refers to the registration method. We actually tested two different 3D non-linear registration methods, FFD and SyN, which will be detailed in the next section.

It is noteworthy that the initialization process of a non-linear registration method is a critical step. On that basis, we used a 3D affine transformation (AFF) based on a *flirt* function using 12 degrees of freedom (available in the FSL toolbox [17]) calculated between each pair of images as an initial step of both FFD and SyN methods. The initialization step for the processing of a $T_1$ weighted image ($T_{1w}$) is defined as: AFF $= T(T_{1w}, T_{1w})_{\mathrm{AFF}}$.

- $\mathrm{FFD}_{T_{1w}}$ method (*ref_f3d* function included in the *NiftiReg* library [21]) is used with the default settings without particular options (final grid spacing ($sx$, $sy$, $sz$) = ($5vx$, $5vx$, $12vx$), number level to perform = 9, maximum iteration = 1,000) and processed from each $T_{1w}$ template included in the source atlas to the $T_{1w}$ target (image to be segmented):

$$\mathrm{FFD}_{T_{1w}} = T(T_{1w}, T_{1w})_{\mathrm{FFD}} \circ \mathrm{AFF}$$

- $\mathrm{SyN}_{T_{1w}}$ method (*ANTS* function included in the *ANTs* library [22]). The parameters were defined to allow a large deformation of areas (cost function = cross-correlation, gradient step = 0.5, total smoothing = 0.5, gradient smoothing = 3, $N$ time steps = 1 and trunk = 256) and processed from each $T_{1w}$ template included in the source atlas to the $T_{1w}$ target:

$$\mathrm{SyN}_{T_{1w}} = T(T_{1w}, T_{1w})_{\mathrm{SyN}} \circ \mathrm{AFF}$$

$\mathrm{SyN}_{\mathrm{ASeg}/T_{1w}}$ method with the same parameters as the $\mathrm{SyN}_{T_{1w}}$ but initialized with the transformation from each ASeg included in the source atlas to the ASeg target, and re-calculated from each $T_{1w}$ registered template included in the source atlas to the $T_{1w}$ target:

$$\mathrm{SyN}_{\mathrm{ASeg}/T_{1w}} = \mathrm{SyN}_{T_{1w}} \circ \mathrm{SyN}_{\mathrm{ASeg}} \circ \mathrm{AFF}$$

The outputs were a set of warp deformation fields. Each deformable field obtained from the registrations was then used in order to resample each MSeg atlas associated with the $T_{1w}$ template image space and to obtain RSeg images (the outputs of registration process, cf. Figs. 3, 4). For each transformation applied to a mask image (MSeg or ASeg), nearest neighbor interpolation was applied to keep the integer values from the original labels.

*Multi-atlas label fusion methods*

The label fusion process was conducted using either the STAPLE or the STEPS algorithm developed by Jorge Cardoso et al. [23], who described it as an extension of the original STAPLE algorithm [24]. The STEPS approach includes a Markov random field aiming at preserving spatial coherence and incorporates a template selection step using a ranking strategy based on the local normalized cross-correlation over a local Gaussian window. The STEPS fusion algorithm implemented in the *NiftiSeg* library was customized and the two main user-defined parameters for the fusion strategy (i.e. the width of the Gaussian kernel for image comparison and the number of labels to use after ranking) were optimized to obtain the highest DSI [25] between MSeg and the results provided by the automatic segmentation pipeline for the four muscles of the QF. This STEPS optimization was performed using both atlases TEST1 and TEST2. For the four muscles of the QF from all the individual sample images and for each combination of parameters (with the number of templates used from 1 to 18 and the Gaussian kernel standard deviation varying from 1 to 12), the average DSI value was found and the higher DSI was used to determine the optimal set of parameters.

Regarding the fusion process, we used the following abbreviations.

$\mathrm{FSeg}^{\mathrm{reg,fus}}_{\mathrm{atlas_{ref}},J} = F(\{\mathrm{RSeg}^{\mathrm{reg}}_{K,J}\})^{\mathrm{fus}}_{K\in\mathrm{atlas_{ref}}}$ the final segmentation of the image $J$ resulting from the fusion $F$, where function and parameters are defined as follows:

$J$: target image, $F$: the transformation describing the fusion, fus: the fusion method (STAPLE or STEPS), $\mathrm{atlas_{ref}}$: the reference multi-atlas, $\{\mathrm{RSeg}^{\mathrm{reg}}_{K,J}\}$: the set of MSeg images $K$ included in the $\mathrm{atlas_{ref}}$ and co-registered to the target image $J$.

*Reliability of muscle volume measurements*

For both atlas-based methods, individual muscle volumes were quantified using the results from MSeg and the truncated cone formula [26] considering the whole set of MRI slices, the slice thickness, and the gaps between slices. Intraclass correlation coefficients (ICC) and coefficients of variation (CV) were calculated using a spreadsheet provided by Hopkins [27] and as previously described [28] in order to assess the reliability of muscle volume measurements and so on, on the basis of a comparative analysis between the results from MSeg and those from the ASeg.

## Results

### Fully automated multi-atlas based segmentation

*Optimization of the STEPS fusion algorithm*

The optimization of the STEPS fusion method relies on the number of templates used for the fusion process and the width of the Gaussian kernel. As illustrated in Fig. 5,

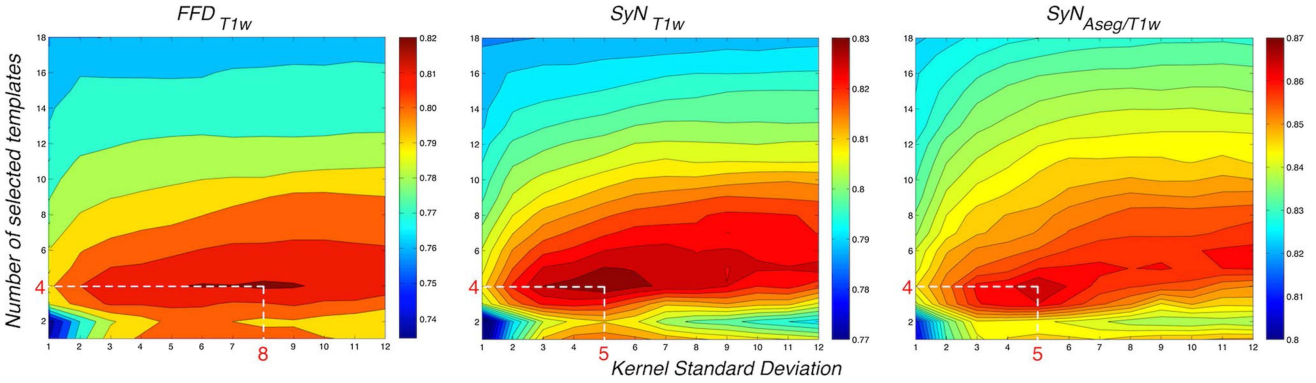


**Fig. 5** Mean Dice similarity index (DSI) for different combinations of number of templates and Gaussian kernel standard deviation used in the STEPS algorithm. The optimal number of templates was the same for FFD and SyN (equal to 4) and the optimal Gaussian kernel standard deviation was equal to 8 for FFD and 5 for SyN (max(DSI) = 0.82 with $FFD_{T_{1w}}$, max(DSI) = 0.84 with $SyN_{T_{1w}}$ and max(DSI) = 0.87 with $SyN_{ASeg/T_{1w}}$ are highlighted with *white dotted lines*)

**Table 1** Metrics for the multi-atlas-based segmentation approach

| Registration method (row) Fusion method (column) | $FFD_{T_{1w}}$ | $SyN_{T_{1w}}$ | $SyN_{ASeg/T_{1w}}$ |
|---|---|---|---|
| STAPLE DSI | 0.76 ± 0.14 | 0.80 ± 0.14 | 0.81 ± 0.14 |
| STEPS DSI | 0.82 ± 0.13 | 0.84 ± 0.13 | 0.87 ± 0.11 |
| STEPS FNVF | 0.13 ± 0.14 | 0.13 ± 0.12 | 0.09 ± 0.10 |
| STEPS FPVF | 0.21 ± 0.15 | 0.17 ± 0.15 | 0.16 ± 0.13 |
| STEPS MVSF | 0.13 ± 0.14 | 0.11 ± 0.14 | 0.13 ± 0.16 |

*DSI* Dice similarity index, *FNVF* false-negative volume fraction, *FPVF* false-positive volume fraction, *MVSF* muscle volume similarity fraction for different registration methods ($FFD_{T_{1w}}$, $SyN_{T_{1w}}$, $SyN_{ASeg/T_{1w}}$) and different fusion algorithms (STAPLE and STEPS) for TEST1 and TEST2 segmented using the MUSCLE25 atlas. Results are presented as mean ± SD

we found the same single local maximum (i.e. a small DSI variation that was closed to the optimal model parameters) for both SyN registration methods corresponding to four templates and a kernel standard deviation ranging between 5 and 8. The corresponding DSI value was 0.82 ± 0.13, 0.84 ± 0.13 and 0.87 ± 0.11 for $FFD_{T_{1w}}$, $SyN_{T_{1w}}$ and $SyN_{ASeg/T_{1w}}$ registration methods, respectively.

### *Comparative analysis of the FFD and SyN registration algorithms combined with STAPLE and STEPS fusion methods*

As the STAPLE algorithm has been largely used in the literature, we compared the corresponding DSI values with those obtained with our pipeline including the STEPS algorithm. The DSI, FNVF, FPVF, and MVSF values obtained for the QF muscle using the STAPLE and STEPS algorithms are presented in Table 1. Regardless of the fusion and the registration methods, the DSI values were systematically larger than 0.7, with the threshold commonly reported as an index of quality. The largest DSI values were systematically obtained for the STEPS algorithm and thus, are considered regardless of the non-linear registration methods. This result was further supported by the box plots displayed in Fig. 6 illustrating a systematic superiority, in terms of DSI values, of the STEPS fusion method in providing higher DSI values and also a narrower range of values. The smallest FNVF (0.09 ± 0.10) and FPVF (0.16 ± 0.13) values were obtained using the $SyN_{ASeg/T_{1w}}$ registration method. However, regarding the accuracy of muscle volume measurements illustrated by the MVSF index, the smallest value (0.11 ± 0.14) was obtained with the $SyN_{T_{1w}}$ registration, although the MVSF values obtained with the other methods were also very low (0.13).

### *Metrics performance of the fully automated multi-atlas-based segmentation and individual muscle volume quantification*

As indicated in Table 2, the average DSI values ranged from 0.72 to 0.94 while the FNVF and the FPVF values ranged from 0.04 to 0.20 and 0.08 to 0.33. Regarding the performance related to the volume measurements, i.e. MVSF, the range was 0.03–0.23. It is noteworthy that the larger volume error (23 %) was observed for both the RF and the VL muscles. Individual muscle volume values were quantified in the QF using the fully automated multi-atlas segmentation pipeline considering the MUSCLE25 atlas as the source and the TEST1 atlases as the target ($FSeg_{(MUSCLE25,TEST1)}$). The corresponding value of the overall QF was 1,129 ± 197 $cm^3$ while individual values ranged from 81 ± 29 $cm^3$ for the RF to 378 ± 64 $cm^3$ for the VM (Table 3). The corresponding averaged value for the QF (1,134 ± 206 $cm^3$) was similar when using TEST2 as the target atlas

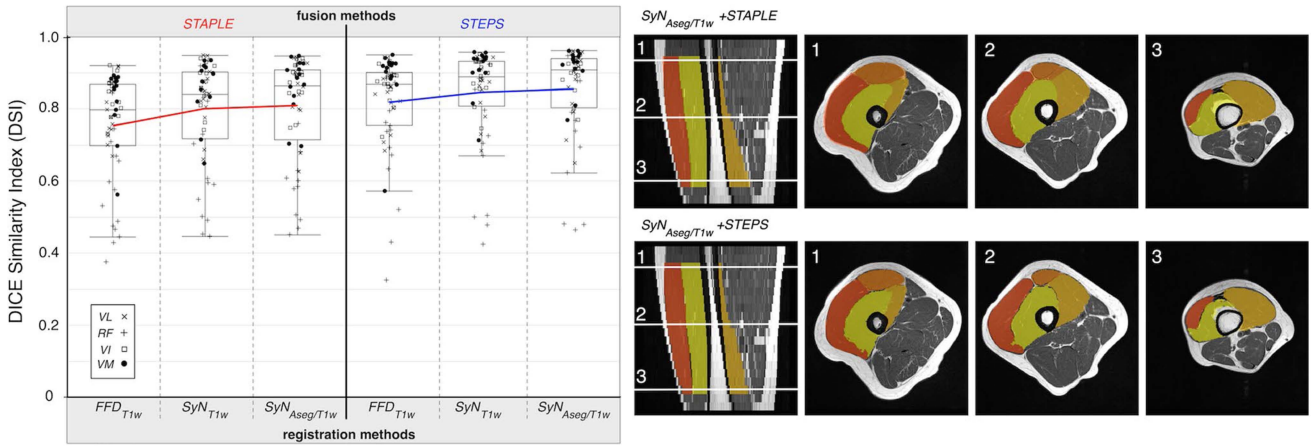


**Fig. 6** *Left panel* box plots showing Dice similarity index (DSI) obtained for the four muscles of the *quadriceps femoris* for both TEST1 and TEST2 atlases. *Right panel* the corresponding segmentation results for STAPLE and optimized STEPS fusion methods are presented for the subject showing the highest Dice similarity index (DSI) value

**Table 2** Metrics related to the multi-atlas based method for each muscle of the *quadriceps femoris*

| Muscle | Manual segmentation | Automated segmentation | DSI | FNVF | FPVF | MVSF |
|---|---|---|---|---|---|---|
| VL | $MSeg_{(TEST1)}$ | $FSeg_{(MUSCLE25,TEST1)}$ | 0.88 ± 0.08 | 0.04 ± 0.02 | 0.17 ± 0.13 | 0.17 ± 0.18 |
| | $MSeg_{(TEST2)}$ | $FSeg_{(MUSCLE25,TEST2)}$ | 0.84 ± 0.10 | 0.05 ± 0.05 | 0.23 ± 0.16 | 0.23 ± 0.23 |
| RF | $MSeg_{(TEST1)}$ | $FSeg_{(MUSCLE25,TEST1)}$ | 0.84 ± 0.12 | 0.06 ± 0.03 | 0.22 ± 0.18 | 0.21 ± 0.24 |
| | $MSeg_{(TEST2)}$ | $FSeg_{(MUSCLE25,TEST2)}$ | 0.72 ± 0.17 | 0.20 ± 0.24 | 0.33 ± 0.13 | 0.23 ± 0.14 |
| VM | $MSeg_{(TEST1)}$ | $FSeg_{(MUSCLE25,TEST1)}$ | 0.94 ± 0.01 | 0.05 ± 0.03 | 0.06 ± 0.01 | 0.03 ± 0.03 |
| | $MSeg_{(TEST2)}$ | $FSeg_{(MUSCLE25,TEST2)}$ | 0.91 ± 0.05 | 0.10 ± 0.06 | 0.08 ± 0.06 | 0.04 ± 0.04 |
| VI | $MSeg_{(TEST1)}$ | $FSeg_{(MUSCLE25,TEST1)}$ | 0.92 ± 0.02 | 0.08 ± 0.05 | 0.08 ± 0.03 | 0.05 ± 0.04 |
| | $MSeg_{(TEST2)}$ | $FSeg_{(MUSCLE25,TEST2)}$ | 0.87 ± 0.07 | 0.12 ± 0.11 | 0.13 ± 0.05 | 0.10 ± 0.07 |

Performance metrics (*DSI* Dice similarity index, *FPVF* false-positive volume fraction, *FNVF* false-negative volume fraction, *MVSF* muscle volume similarity fraction values) of the multi-atlas segmentation approach for the individual muscles of the *quadriceps femoris* (*VL*, *vastus lateralis*; *RF*, *rectus femoris*; *VM*, *vastus medialis*; *VI*, *vastus intermedius*). Metrics have been calculated on the basis of a comparative analysis between the automated segmentation (FSeg) and manual segmentations (MSeg) of TEST1 and TEST2 atlases, using $SyN_{ASeg/T_{1w}}$ registration method and STEPS fusion. Values are presented as mean ± SD

**Table 3** Quantitative volume measurements for each muscle and for each segmentation method

| Volumes ($cm^3$) | MSeg | | FSeg | | RSeg | |
|---|---|---|---|---|---|---|
| Muscle | TEST1 | TEST2 | (MUSCLE25,TEST1) | (MUSCLE25,TEST2) | (TEST2,TEST1) | (TEST1,TEST2) |
| VL | 286 ± 111 | 286 ± 112 | 322 ± 82 | 340 ± 88 | 283 ± 112 | 295 ± 115 |
| RF | 69 ± 35 | 67 ± 32 | 81 ± 29 | 72 ± 18 | 67 ± 31 | 71 ± 37 |
| VM | 375 ± 70 | 378 ± 68 | 378 ± 64 | 371 ± 67 | 385 ± 78 | 376 ± 68 |
| VI | 349 ± 51 | 351 ± 53 | 348 ± 53 | 351 ± 55 | 350 ± 51 | 355 ± 54 |
| QF | 1080 ± 220 | 1082 ± 229 | 1129 ± 197 | 1134 ± 206 | 1085 ± 229 | 1098 ± 231 |

Individual quantitative volumes are estimated for each muscle (*VL*, *vastus lateralis*; *RF*, *rectus femoris*; *VM*, *vastus medialis*; *VI*, *vastus intermedius* of the *QF*, *quadriceps femoris*) and for each segmentation (*MSeg* manual segmentation, *FSeg* automated multi-atlas-based segmentation, *RSeg* semi-automated single-atlas-based segmentation) of TEST1 and TEST2 atlases. The mean ± SD volumes are measured in $cm^3$

($FSeg_{(MUSCLE25,TEST2)}$). Individual muscle volumes values ranged from 72 ± 18 $cm^3$ for the RF to 371 ± 67 $cm^3$ for the VM (Table 3). Generally speaking, the ground-truth results, i.e. resulting from the manual segmentation ($MSeg_{(TEST1)}$ and $MSeg_{(TEST2)}$) were systematically lower, e.g., 1,080 ± 220 and 1,082 ± 229 $cm^3$ for the QF using TEST1 and TEST2 atlases, respectively (Table 3). Individual values were reported in supplementary Table 1.

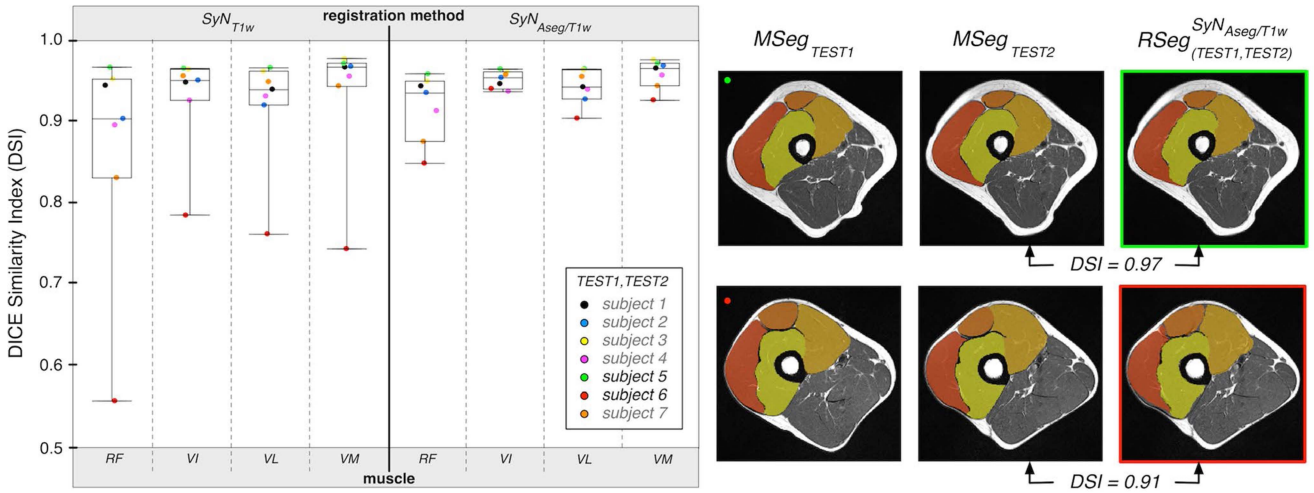


**Fig. 7** *Left panel* box plots showing DSI values obtained for the four muscles of the *quadriceps femoris* using $SyN_{T_{1w}}$ and $SyN_{ASeg/T_{1w}}$ registration methods from the whole subjects included in the TEST1 atlas to the corresponding subjects in the TEST2 atlas (*each colour* represents an individual subject). *Right panel* the corresponding segmentation results for the $SyN_{ASeg/T_{1w}}$ registration method are presented for two subjects showing the highest (*top*) and the lowest (*bottom*). Dice similarity index (DSI) values calculated on the basis of a comparative analysis between the manual segmentation ($MSeg_{(TEST2)}$) and the semi-automated single-atlas-based segmentation ($RSeg_{(TEST1,TEST2)}$)

## Semi-automated single-atlas-based segmentation

### *Evaluation of the SyN registration algorithms*

In order to illustrate the potential of the various non-linear registration methods, we assessed both the $SyN_{T_{1w}}$ and the $SyN_{ASeg/T_{1w}}$ methods for repeated measurements in seven healthy subjects (TEST1 co-registered on the corresponding images of the TEST2 atlas and vice versa). This approach is similar to a more classic individual single-atlas-based non-linear registration method in the context of longitudinal tracking.

As illustrated in Fig. 7, the corresponding average DSI value for the QF was 0.91 ± 0.09 for $SyN_{T_{1w}}$ whereas it was higher (0.94 ± 0.03) for the $SyN_{ASeg/T_{1w}}$, thereby illustrating the key importance of the initialization process of the non-linear registration method. In addition, to the overall increase of the DSI values, the $SyN_{ASeg}$ initialization process led to a substantial reduction of the DSI values range (Fig. 7). This was particularly more the case for the RF muscle, which is the smallest muscle within the QF and also displayed the largest morphology variability in our population.

### *Metrics performance of the semi-automated single-atlas-based segmentation method and individual muscle volume quantification*

As indicated in Table 4, the average DSI values ranged from 0.89 to 0.95 while the FNVF and the FPVF fractions values ranged from 0.04 to 0.10. Regarding the performance related to the volume measurements MVSF, the range was 0.02–0.05. The larger volume error (5 %) was observed for the RF muscle.

Individual muscle volume values were quantified in the QF using the semi-automated single-atlas segmentation pipeline considering either TEST1 or TEST2 as source or target atlases and vice versa. Using TEST1 as the target atlas ($RSeg_{(TEST2,TEST1)}$), the QF volume was 1,085 ± 229 $cm^3$ and was almost identical (1,098 ± 231 $cm^3$) using TEST2 as the target atlas ($RSeg_{(TEST1,TEST2)}$). Similarly, individual muscle volumes ranged from 67 ± 31 $cm^3$ for the RF to 385 ± 78 $cm^3$ for the VM using TEST1 as the target atlas and from 71 ± 37 $cm^3$ for the RF to 376 ± 68 $cm^3$ for the VM (Table 3).

## Reproducibility measurements of quantitative volume using both atlas-based approaches

The reproducibility of each atlas-based method was assessed on the basis of the ICC and CV measurements from a comparative analysis between the manual and the automated volume measurements on both TEST1 and TEST2 atlases. The corresponding values are reported in Table 5, for each muscle and for different combinations. The ground-truth values were computed from a comparative analysis between repeated manual volume measurements; the ICC values were larger than 0.98 and the corresponding volume measurement error (illustrated by the CV) was 2.0 % for the whole QF muscle and ranged from 1.7 (VI) to 5.5 % (VL). As indicated in Table 5, the

**Table 4** Metrics related to the single-atlas-based method for each muscle of the *quadriceps femoris*

| Muscle | Manual segmentation | Automated segmentation | DSI | FNVF | FPVF | MVSF |
|---|---|---|---|---|---|---|
| VL | $MSeg_{(TEST1)}$ | $RSeg_{(TEST2,TEST1)}$ | 0.94 ± 0.02 | 0.07 ± 0.03 | 0.04 ± 0.01 | 0.03 ± 0.03 |
| | $MSeg_{(TEST2)}$ | $RSeg_{(TEST1,TEST2)}$ | 0.93 ± 0.03 | 0.06 ± 0.02 | 0.07 ± 0.05 | 0.03 ± 0.04 |
| RF | $MSeg_{(TEST1)}$ | $RSeg_{(TEST2,TEST1)}$ | 0.92 ± 0.04 | 0.10 ± 0.05 | 0.06 ± 0.03 | 0.05 ± 0.04 |
| | $MSeg_{(TEST2)}$ | $RSeg_{(TEST1,TEST2)}$ | 0.89 ± 0.07 | 0.10 ± 0.06 | 0.10 ± 0.09 | 0.05 ± 0.03 |
| VM | $MSeg_{(TEST1)}$ | $RSeg_{(TEST2,TEST1)}$ | 0.96 ± 0.02 | 0.04 ± 0.01 | 0.04 ± 0.03 | 0.02 ± 0.02 |
| | $MSeg_{(TEST2)}$ | $RSeg_{(TEST1,TEST2)}$ | 0.95 ± 0.04 | 0.06 ± 0.05 | 0.04 ± 0.04 | 0.02 ± 0.03 |
| VI | $MSeg_{(TEST1)}$ | $RSeg_{(TEST2,TEST1)}$ | 0.95 ± 0.01 | 0.05 ± 0.02 | 0.04 ± 0.01 | 0.02 ± 0.02 |
| | $MSeg_{(TEST2)}$ | $RSeg_{(TEST1,TEST2)}$ | 0.94 ± 0.01 | 0.05 ± 0.01 | 0.05 ± 0.02 | 0.02 ± 0.01 |

Performance metrics (*DSI* Dice similarity index, *FPVF* false-positive volume fraction, *FNVF* false-negative volume fraction, *MVSF* muscle volume similarity fraction values) of the single-atlas segmentation approach for the individual muscles of the *quadriceps femoris* (*VL*, *vastus lateralis*; *RF*, *rectus femoris*; *VM*, *vastus medialis*; *VI*, *vastus intermedius*). Metrics have been calculated on the basis of a comparative analysis between the automated segmentations (RSeg) and manual segmentations (MSeg) of TEST1 and TEST2 atlases, using the $SyN_{ASeg/T_{1w}}$ registration method. Values are presented as mean ± SD

**Table 5** Reproducibility of volume measurements for each muscle and for each atlas-based method

| Muscle | $MSeg_{(TEST1)}$ $MSeg_{(TEST2)}$ | | $MSeg_{(TEST1)}$ $FSeg_{(MUSCLE25,TEST1)}$ | | $MSeg_{(TEST2)}$ $FSeg_{(MUSCLE25,TEST2)}$ | | $MSeg_{(TEST1)}$ $RSeg_{(TEST2,TEST1)}$ | | $MSeg_{(TEST2)}$ $RSeg_{(TEST1,TEST2)}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ICC | CV (%) | ICC | CV (%) | ICC | CV (%) | ICC | CV (%) | ICC | CV (%) |
| VL | 0.99 | 5.5 | 0.94 | 15.7 | 0.90 | 20.6 | 0.99 | 3.2 | 1.00 | 2.1 |
| RF | 0.98 | 4.5 | 0.88 | 22.4 | 0.78 | 17.3 | 0.99 | 4.0 | 0.99 | 3.6 |
| VM | 0.99 | 2.1 | 0.98 | 3.2 | 0.96 | 3.9 | 0.98 | 1.1 | 0.99 | 1.8 |
| VI | 0.98 | 1.7 | 0.90 | 5.3 | 0.70 | 10.5 | 0.98 | 2.0 | 0.99 | 1.7 |
| QF | 0.99 | 2 | 0.98 | 3.2 | 0.97 | 4.4 | 0.99 | 1.4 | 1.00 | 0.9 |

Intraclass correlation coefficient (ICC) and coefficient of variation (CV) calculated for each individual muscle volume (*VL*, *vastus lateralis*; *RF*, *rectus femoris*; *VM*, *vastus medialis*; *VI*, *vastus intermedius* of the *QF*, *quadriceps femoris*) and for each atlas-based method [semi-automated, single-atlas-based segmentation (RSeg) and automated, multi-atlas-based segmentation (FSeg)]. For each method the manual segmentation (MSeg) results were considered as the ground truth. Measurements were performed for both TEST1 and the TEST2 atlases using the $SyN_{ASeg/T_{1w}}$ registration method (for both approaches) and the STEPS fusion (for multi-atlas only)

performance of the single-atlas method was actually similar to the ground-truth method with an ICC value larger than 0.99 and a volume error smaller than 4.0 %. On the contrary, the multi-atlas method performance was much worse. The results might be considered as acceptable for the whole QF muscle (ICC larger than 0.97 and CV smaller than 4.4 %). However, the corresponding values for the individual muscles were much larger, ranging from 3.2 to 22.4 % for the CV and from 0.70 to 0.98 for the ICC.

## Discussion

In the present study we assessed the performance of two atlas-based methods using non-linear registration processes with the aims of providing an automatic segmentation of individual muscles and quantifying the corresponding muscle volumes. Previous studies have clearly indicated that manual labelling of anatomical structures was time consuming and expert-dependent such that automated techniques would be of interest [1, 5, 6]. Although a few studies have assessed the potential of multi-atlas-based methods for muscle segmentation [14], individual muscle volume measurements have scarcely been reported and the performance of the different methods has been mainly assessed on the basis of the DSI, which can be questioned for muscle volume measurements.

On the basis of several metrics regarding muscle segmentation, we clearly showed that the fully automated, multi-atlas-based segmentation was robust and that the semi-automated, single-atlas segmentation provided more robust and accurate results. Regarding muscle volume measurements, the multi-atlas-based method was acceptable for the overall QF muscle with a 3.2–4.4 % measurement error. However, the results regarding individual muscles were not acceptable, with measurement errors ranging from 3.2 to 22.4 %. On that basis, the very low measurement error associated to the single-atlas method (1.1–4.0 %) clearly positioned this method as the method of choice for longitudinal measurements for which subtle volume changes might have to be quantified.

### Non-registration and fusion processes

For both non-linear registration methods, i.e. FFD and SyN, we observed a large between-muscle variability for the DSI values using the STAPLE algorithm. Using the STEPS method, this variability was substantially reduced and the individual DSI values were increased such that this fusion method was chosen. The fusion method strategy allowed us to only select appropriate templates for the segmentation process and we showed that the selection of four templates was optimal for the MUSCLE25 atlas. The necessity of such a step has been clearly identified in the study of Prescott et al. [5], in which they used, based on normalized histograms, a single template among the six available in order to increase the accuracy of the segmentation process.

### Metrics of the atlas-based approaches

On the basis of a comparative analysis between the results from the manual and the fully automated, multi-atlas-based segmentation, we clearly demonstrated that our segmentation approach of the four muscles of the QF was reliable. Regarding the DSI values, the performance of our multi-atlas-based method combining a non-linear registration method to a fusion process was rather high. As a matter of comparison, a value of 0.7 has been reported as a good agreement in the context of morphometric analysis of white matter lesions in brain [15]. The average DSI value ($0.87 \pm 0.11$) we reported for the $\mathrm{SyN}_{\mathrm{ASeg}/T_{1w}}$ method was slightly higher than the values obtained using a single-atlas-based level set segmentation strategy (RF $0.78 \pm 0.12$; VI $0.79 \pm 0.10$; VL $0.82 \pm 0.08$ and VM $0.69 \pm 0.16$) [5] and comparable to those obtained using a random walks-based approach for individual muscles ($0.86 \pm 0.07$) [3]. It is noteworthy that our DSI results were also similar to those reported for the segmentation of prostate (DSI = 0.85) [11] and the pectoral muscle (DSI = 0.74) [12]. Of interest, the robustness of the semi-automated, individual, single-atlas approach was systematically higher on the basis of the DSI values, which were always larger than 0.9.

### Volume measurements using the atlas-based approaches

The other metrics reported in the present study allowed us to evaluate the accuracy of the muscle volume measurements. Considering that these metrics have been very scarcely reported in other studies intending to use atlas-based segmentation for the quantification of muscle volumes, we were only able to compare these values with a single study. The FNVF, FPVF and MVSF values we reported in the present study were very similar to those reported by Karlsson et al. [14] for the whole QF muscle. However, the situation was very different for the volume quantification of individual muscles. While the single-atlas approach provided volume measurements errors (1.1–4.0 % error) very similar to those resulting from the ground-truth analysis (1.7–5.5 % volume error), i.e. the repeated MSegs, the multi-atlas-based method provided larger errors (3.2–22.4 %), clearly indicating that subtle changes in individual muscle volumes could not be accurately quantified using a multi-atlas-based strategy and that a single-atlas-based approach must be preferred. This has to be considered with respect to previous longitudinal investigations conducted in healthy subjects and indicating a selective hypertrophy among or along the whole QF after a resistance-training program [29, 30], an acute exercise [31], or a muscle-specific atrophy after a short- and long-term immobilization period [32, 33].

For both segmentation methods, we found that the robustness differed between individual muscles with a higher performance systematically observed for the VI and VM muscles and a poorer performance for the RF and VL muscles. Such an anatomical dependence has been already reported in previous studies. While the VM muscle was reported as the most variable muscle by Prescott et al. [5], Karlsson et al. [14] reported a larger variability in the upper part of the body as compared to the lower part. In the latter study, it has been suggested that this variability was not only due to purely anatomical variations but also to the subject positioning in the MRI scanner, thereby indicating the importance of the subject positioning for longitudinal and also for cross-sectional studies. The multi-atlas dataset we have constructed relied on MR images recorded in a very straightforward way regarding the leg position, and this certainly substantially improved the accuracy of the segmentation approach.

### Usefulness of atlas-based approaches for longitudinal tracking

At this stage, one could question the usefulness of these methods for the longitudinal tracking of pathological changes occurring in patients with neuromuscular disorders. Previous longitudinal studies have been conducted in patients but the corresponding analyses have been mainly qualitative [34, 35]. The segmentation strategies reported so far have never tackled the intermuscular fatty infiltration, which is one of the main hallmarks of the disease in these patients and is likely to challenge the robustness of any segmentation strategy. Although additional studies would have to further investigate this issue, it looks as if the multi-atlas-based strategy is inappropriate given that a relevant multi-atlas database would have to be available for each disease and for each stage of a disease. The single-atlas strategy might be a relevant alternative inasmuch as longitudinal

measurements of changes in muscle volume and fat infiltration are of interest. We emphasize that the larger performance of the single-atlas method is counterbalanced by a higher processing time as compared to the multi-atlas-based approach. However, one has to keep in mind that in a context of multiple MRI performed for a given subject, the MSeg process would have to be performed only once and used as an atlas for the segmentation of the following MRI datasets. According to our results, this is the optimal choice if one aims at measuring individual muscle volume and performs a longitudinal follow-up for a given subject. For individual muscles, we reported a very high accuracy of the single-atlas method on the basis of measurements repeated over a 2-day period, and one can wonder what the situation could be for a 2-year period. Based on the results from Maden-Wilkinson et al. [36] showing a low average muscle volume loss (0.77 % per year), the reproducibility we reported for the 2-day period should be similar to what we could have measured over a 2-year-period. For patients with neuromuscular disorders, both inter- and intramuscular fatty infiltrations might challenge the segmentation process and this issue would have to be investigated in future studies.

### Challenging factors related to atlas-based approaches

Additional challenging factors such as MRI artifacts and intramuscular fat have also been identified for muscles segmentation strategies [5, 37]. For the limited MRI dataset we used for the present study, we did not identify any particular artifact that could challenge the quality of the ASeg. Our strategy of combining the non-linear registration with a fusion method did not suffer from the presence of intramuscular fat. Age, sex, and body mass index have also been reported as potential challenging factors for an ASeg strategy given that these factors provide additional variability regarding muscle shape, volume, SAT thickness, and the degree of IMAT infiltration [37]. The present study was conducted in healthy men ranging from 25 to 40 years and we only observed a slight variability among subjects and muscles. The extension of the present database with women and older subjects should allow us to take into account the natural diversity related to age and sex, as previously described [13]. However, it should be mentioned that challenging factors are likely to be related to the segmentation strategy. While Karlsson et al. [14] mentioned that age and BMI did not affect the performance of their methods, Prescott et al. [5] indicated that the poorer performance of their algorithm for the VM muscle could be due to the larger variability of the muscle shape and the limited consistency of the corresponding borders.

The present work is a prospective study in the growing field of automated muscle segmentation and has some limitations related to the limited number of subjects, although comparable to the populations size used in previous studies [3, 5, 14]. This approach should be further tested in a larger cohort of healthy subjects.

## Conclusion

We have presented two atlas-based methods allowing a segmentation of individual muscles from the QF group and the corresponding volume quantification. Both segmentation algorithms were very robust and the performances related to muscle volume measurements were very different. While the fully automated, multi-atlas-based approach was satisfactory for global volume measurement, the semi-automated, single-atlas-based approach must be preferred for volume measurements of individual muscles and for longitudinal investigations.

**Acknowledgments** This study was supported by the Centre National de la Recherche Scientifique (CNRS, UMR 7339). The authors thank the subjects who participated in the present study.

**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

1. Barnouin Y, Butler-Browne G, Voit T, Reversat D, Azzabou N, Leroux G, Behin A, McPhee JS, Carlier PG, Hogrel J-Y (2014) Manual segmentation of individual muscles of the quadriceps femoris using MRI: A reappraisal. J Magn Reson Imaging 40:239–247
2. Baudin PY, Azzabou N, Carlier PG, Paragios N (2012) Prior knowledge, random walks and human skeletal muscle segmentation. Med Image Comput Comput Assist Interv 15:569–576
3. Baudin P-Y, Goodman D, Kumrnar P, Azzabou N, Carlier PG, Paragios N, Kumar MP (2013) Discriminative parameter estimation for random walks segmentation. Med Image Comput Comput Assist Interv 16:219–226
4. Makrogiannis S, Serai S, Fishbein KW, Schreiber C, Ferrucci L, Spencer RG (2012) Automated quantification of muscle and fat in the thigh from water-, fat-, and nonsuppressed MR images. J Magn Reson Imaging 35:1152–1161
5. Prescott JW, Best TM, Swanson MS, Haq F, Jackson RD, Gurcan MN (2011) Anatomically anchored template-based level set segmentation: Application to quadriceps muscles in MR images from the osteoarthritis initiative. J Digit Imaging 24:28–43

6. Ahmad E, Yap MH, Degens H, McPhee JS (2014) Atlas-registration based image segmentation of MRI human thigh muscles in 3D space. In: Proceedings of SPIE 9037. Medical imaging 2014: image perception, observer performance, and technology assessment. doi:10.1117/12.2043606
7. Ma D, Cardoso MJ, Modat M, Powell N, Wells J, Holmes H, Wiseman F, Tybulewicz V, Fisher E, Lythgoe MF, Ourselin S (2014) Automatic structural parcellation of mouse brain MRI using multi-atlas label fusion. PLoS One 9:e86576
8. Eugenio Iglesias J, Rory Sabuncu M, Van Leemput K (2013) A unified framework for cross-modality multi-atlas segmentation of brain MRI. Med Image Anal 17:1181–1191
9. Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A (2006) Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33:115–126
10. Wu G, Wang Q, Zhang D, Nie F, Huang H, Shen D (2014) A generative probability model of joint label fusion for multi-atlas based brain segmentation. Med Image Anal 18:881–890
11. Klein S, van der Heide UA, Lips IM, van Vulpen M, Staring M, Pluim JPW (2008) Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med Phys 35:1407
12. Gubern-Mérida A, Kallenberg M, Martí R, Karssemeijer N (2012) Segmentation of the pectoral muscle in breast MRI using atlas-based approaches. Med Image Comput Comput Assist Interv 15:371–378
13. Thomas MS, Newman D, Leinhard OD, Kasmai B, Greenwood R, Malcolm PN, Karlsson A, Rosander J, Borga M, Toms AP (2014) Test–retest reliability of automated whole body and compartmental muscle volume measurements on a wide bore 3T MR system. Eur Radiol 24:2279–2291
14. Karlsson A, Rosander J, Romu T, Tallberg J, Grönqvist A, Borga M, Dahlqvist Leinhard O (2015) Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI. J Magn Reson Imaging 41:1558–1569
15. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC (1994) Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Trans Med Imaging 13:716–724
16. Udupa JK, LeBlanc VR, Zhuge Y, Imielinska C, Schmidt H, Currie LM, Hirsch BE, Woodburn J (2006) A framework for evaluating image segmentation algorithms. Comput Med Imaging Graph 30:75–87
17. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790
18. Positano V, Christiansen T, Santarelli MF, Ringgaard S, Landini L, Gastaldelli A (2009) Accurate segmentation of subcutaneous and intermuscular adipose tissue from MR images of the thigh. J Magn Reson Imaging 29:677–684
19. Xu C, Prince JL (1998) Snakes, shapes, and gradient vector flow. IEEE Trans Image Process 7:359–369
20. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans Med Imaging 18:712–721
21. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S (2010) Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed 98:278–284
22. Avants B, Epstein C, Grossman M, Gee J (2008) Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12:26–41
23. Jorge Cardoso M, Leung K, Modat M, Keihaninejad S, Cash D, Barnes J, Fox NC, Ourselin S (2013) STEPS: similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation. Med Image Anal 17:671–684
24. Warfield SK, Zou KH, Wells WM (2004) Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23:903–921
25. Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, Wells WM, Jolesz FA, Kikinis R (2004) Statistical validation of image segmentation quality based on a spatial overlap index 1: Scientific reports. Acad Radiol 11:178–189
26. Nordez A, Jolivet E, Südhoff I, Bonneau D, de Guise JA, Skalli W (2009) Comparison of methods to assess quadriceps muscle volume using magnetic resonance imaging. J Magn Reson Imaging 30:1116–1123
27. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med Auckl N Z 30:1–15
28. Shrout PE, Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 86:420–428
29. Reeves ND, Narici MV, Maganaris CN (2004) Effect of resistance training on skeletal muscle-specific force in elderly humans. J Appl Physiol (1985) 96:885–892
30. Gondin J, Guette M, Ballay Y, Martin A (2005) Electromyostimulation training effects on neural drive and muscle architecture. Med Sci Sports Exerc 37:1291–1299
31. Fouré A, Duhamel G, Wegrzyk J, Boudinet H, Mattei J-P, Le Troter A, Bendahan D, Gondin J (2015) Heterogeneity of muscle damage induced by electrostimulation: A multimodal MRI study. Med Sci Sports Exerc 47:166–175
32. Thom JM, Thompson MW, Ruell PA, Bryant GJ, Fonda JS, Harmer AR, Janse de Jonge XA, Hunter SK (2001) Effect of 10-day cast immobilization on sarcoplasmic reticulum calcium regulation in humans. Acta Physiol Scand 172:141–147
33. Vandenborne K, Elliott MA, Walter GA, Abdus S, Okereke E, Shaffer M, Tahernia D, Esterhai JL (1998) Longitudinal study of skeletal muscle adaptations during immobilization and rehabilitation. Muscle Nerve 21:1006–1012
34. Willis TA, Hollingsworth KG, Coombs A, Sveen M-L, Andersen S, Stojkovic T, Eagle M, Mayhew A, de Sousa PL, Dewar L, Morrow JM, Sinclair CDJ, Thornton JS, Bushby K, Lochmuller H, Hanna MG, Hogrel J-Y, Carlier PG, Vissing J, Straub V (2014) Quantitative magnetic resonance imaging in limb-girdle muscular dystrophy 2I: A multinational cross-sectional study. PLoS one 9:e90377
35. Hollingsworth KG, Garrood P, Eagle M, Bushby K, Straub V (2013) Magnetic resonance imaging in Duchenne muscular dystrophy: Longitudinal assessment of natural history over 18 months: short reports. Muscle Nerve 48:586–588
36. Maden-Wilkinson TM, Degens H, Jones DA, McPhee JS (2013) Comparison of MRI and DXA to measure muscle size and age-related atrophy in thigh muscles. J Musculoskelet Neuronal Interact 13:320–328
37. Hogrel J-Y, Barnouin Y, Azzabou N, Butler-Browne G, Voit T, Moraux A, Leroux G, Behin A, McPhee JS, Carlier PG (2015) NMR imaging estimates of muscle volume and intramuscular fat infiltration in the thigh: Variations with muscle, gender, and age. Age. doi:10.1007/s11357-015-9798-5