Automated Measurement of Fat Infiltration in the Hip Abductors from Dixon Magnetic Resonance Imaging

Martin A. Belzunce, PhD[1], Johann Henckel, MD[1], Anastasia Fotiadou, PhD,[1], Anna Di Laura, PhD[1], and Alister Hart, MD[1,2]

1 Royal National Orthopaedic Hospital, Stanmore HA7 4LP, UK

2 Institute of Orthopaedics and Musculoskeletal Science, University College London, Stanmore HA7 4LP, UK

Corresponding author: Martin A. Belzunce at Research, Royal National Orthopaedic Hospital (RNOH), Brockley Hill, Stanmore, Middlesex, HA7 4LP. E-mail: martin.belzunce@nhs.net.

Abstract

Purpose

Intramuscular fat infiltration is a dynamic process, in response to exercise and muscle health, which can be quantified by estimating fat fraction (FF) from Dixon MRI. Healthy hip abductor muscles are a good indicator of a healthy hip and an active lifestyle as they have a fundamental role in walking. The automated measurement of the abductors' FF requires the challenging task of segmenting them. We aimed to design, develop and evaluate a multi-atlas based method for automated measurement of fat fraction in the main hip abductor muscles: gluteus maximus (GMAX), gluteus medius (GMED), gluteus minimus (GMIN) and tensor fasciae latae (TFL).

Method

We collected and manually segmented Dixon MR images of 10 healthy individuals and 7 patients who underwent MRI for hip problems. Twelve of them were selected to build an atlas library used to implement the automated multi-atlas segmentation method. We compared the FF in the hip abductor muscles for the automated and manual segmentations for both healthy and patients groups. Measures of average and spread were reported for FF for both methods. We used the root mean square error (RMSE) to quantify the method accuracy. A linear regression model was used to explain the relationship between FF for automated and manual segmentations.

Results

The automated median (IQR) FF was 20.0(16.0-26.4) %, 14.3(10.9-16.5) %, 15.5(13.9-18.6) % and 16.2(13.5-25.6) % for GMAX, GMED, GMIN and TFL respectively, with a FF RMSE of 1.6%, 0.8%, 2.1%, 2.7%. A strong linear correlation (R^2=0.93, p<0.001, m=0.99) was found

between the FF from automated and manual segmentations. The mean FF was higher in patients than in healthy subjects.

Conclusion

The automated measurement of FF of hip abductor muscles from Dixon MRI had good agreement with FF measurements from manually segmented images. The method was accurate for both healthy and patients groups.

Keywords: hip abductors, multi-atlas segmentation, Dixon, fat fraction, muscle segmentation; fat infiltration

Automated Measurement of Fat Infiltration in the Hip Abductors from Dixon Magnetic Resonance Imaging

1. Introduction

The hip abductor muscles have a fundamental role in running, walking, standing and other human daily activities (1, 2), and can be considered a good indicator of a healthy hip and an active lifestyle.

An increase of intermuscular adipose tissue (IMAT) and intramuscular fat (IMF) is associated with loss of strength and mobility dysfunction (3), making it an important marker for muscle health. Fat infiltration in skeletal muscles is linked to aging/sarcopenia (4–6), orthopaedic conditions (7, 8), muscular dystrophies (9, 10) and physiological disorders (11–13); and can be observed and quantified with magnetic resonance imaging (MRI) (6, 10, 12), a more objective assessment method than the commonly used manual functional tests.

Fat-water separation techniques, such as Dixon MR imaging (14–17), provide fat-only and water-only images which can be used to estimate a fat fraction (FF) image that quantifies fat content. FF from Dixon images has been successfully used to assess fat infiltration in the thigh muscles in different scenarios (11, 13, 18). The measurement of FF in individual muscles requires the labelling or segmentation of them, which is a difficult task to automate as individual muscles share similar intensity/contrast values.

Multi-atlas methods have proved to be the most successful method for automatically segmenting the thigh muscles (19–23). In the case of the hip abductors, this is a more challenging task due to its more complex and heterogeneous anatomy, and only a small number of semi-automated (24, 25) and automated methods (26, 27) have been presented. To date, multi-atlas methods have not shown to be accurate enough to be able to measure small and medium differences in muscle volume. However, the segmentation accuracy required for automated measurement of FF is lower than for volume, since a mean intensity value within a label is measured. In addition, changes in muscle composition are proportionally larger than changes in volume, and a better predictor for mobility and muscle strength than volume changes (3, 28–31). We hypothesize that the combination of an optimized multi-atlas segmentation method and Dixon MRI of the pelvis can achieve accurate FF measurements of the abductor muscles, and that this accuracy is sufficient to study cross-sectional differences in muscle composition.

We aimed to evaluate an automated method for measurement of fat fraction in the main hip abductor muscles: gluteus maximus (GMAX), gluteus medius (GMED), gluteus minimus (GMIN) and tensor fasciae latae (TFL). To do this we 1) collected and manually segmented Dixon MR images of 10 healthy individuals and 7 patients who underwent MRI for hip problems; 2) designed and implemented a multi-atlas based method for segmentation of the hip abductor muscles; 3) measured the FF in the muscles for both manually and automatically segmented Dixon images; 4) evaluated the performance of the automated method using the FF values from the manual segmentations as a reference; and 5) compared the FF results in both healthy and patient groups for both methods.

2. Materials and Methods

In this work, we designed, developed and evaluated an automated method to measure fat fraction in the hip abductor muscles. In Figure 1, a flowchart describing the procedure to implement and evaluate our method is shown. Dixon MRI scans from 10 healthy subjects and 7 patients with diseased hips were collected and GMAX, GMED, GMIN and TFL were manually segmented in the Dixon images. We used scans from subjects with both healthy and diseased hips in order to evaluate our method in different scenarios and with a higher variability of fat infiltration. Twelve out of the 17 segmented images were used to create an atlas library, which was employed in our automated multi-atlas based method for the segmentation of the hip abductor muscles. The latter was used to segment the 17 Dixon scans. Finally, the FF in each muscle was computed for the automatically and manually segmented images. The accuracy of the automated FF measurements were evaluated using the FF values from the manual segmentations as a reference. The segmentation performance was assessed using the segmented images.

2.1 Study Subjects and Data Acquisition

We studied Dixon MRI scans from a group of healthy volunteers (HV) and a group of patients with diseased hips. The group of healthy subjects consisted of 10 subjects recruited for a study looking at the effects of marathon running in the hip joints. This group went through a standardized MRI protocol, including a Dixon scan of the full pelvis. For the patients' group, we retrospectively collected 7 scans of patients with OA or other hip conditions that had a full pelvis Dixon MRI scan in the past. The demographic characteristics of the two groups are shown in Table 1. All subjects consented to the study.

The MR images for the HV group were acquired on a 3T scanner (Siemens Magneton Vida, Erlangen, Germany) using a body coil. The scanning protocol consisted of standard clinical sequences for the hips; axial Dixon (slice thickness 1.5 mm, spacing between slices 1.95 mm, repetition time (TR) 4570 msec, echo time (TE) 45 msec, number of excitations 1, number of echoes 14, flip angle 120°) and axial T1-weighted turbo spin echo (slice thickness 3.0 mm, spacing between slices 3.3 mm, TR 895 msec, TE 8.9 msec) sequences of the pelvis. The Dixon sequence was especially designed to assess gluteal muscles and had a field of view (FOV) that covered axially from 3 cm below the lesser trochanter to the top of the iliac crest. The voxel size was $0.47\times0.47\times1.95$ mm^3.

The Dixon sequence in the patients' scans had the same parameters used for the healthy volunteers group, except for three cases that were scanned with lower resolution to accommodate the sequence in the restricted acquisition time. In the latter, the voxel size was $1.19\times1.19\times3.3$ mm^3.

2.2 Fat Fraction in the Hip Abductors

We computed the FF, as a measure of fat infiltration, in each hip abductor muscle using the fat and water images of the Dixon sequence. The FF of each abductor is defined as:

$$FF_l[\%] = \frac{1}{N_l}\sum_{i\,\epsilon\,S_l}\frac{F_i}{W_i+F_i}100\% \qquad (1)$$

where FF_l is the fat fraction in the labelled muscle l, i is the index of each of the N_l voxels in the set S_l that represent the label l; and F_i and W_i are the values of voxel i in the fat and water images respectively.

In the proposed method, the labels for each muscle are obtained with a fully automated multi-atlas segmentation algorithm. Next, a mask is generated from each label and is eroded with a spherical structuring element of radius 1 voxel to avoid the muscle edges. Finally, the FF in each muscle is computed using the eroded mask and equation (1). The FF from manually segmented images were also included to validate the method.

A graphical example of the estimation of the FF in the right GMAX from a Dixon scan is shown in Figure 2. The in-phase image is used in the segmentation, while the water and fat images in the FF calculation.

2.3 Manual Segmentation

A single experienced operator labelled GMAX, GMED, GMIN and TFL muscles on the in-phase image of every scan. The out-of-phase image was also used to aid the labelling process. A different label was used for left and right muscles, rounding up 8 labels per atlas. IMAT was excluded while IMF was included as part of each muscle. The segmentation was carried out in Simpleware™ ScanIP (Version 2018.12; Synopsys, Inc., Mountain View, USA), an FDA and CE marked 3D image processing software for medical scan data.

2.4 Automated Multi-Atlas Segmentation

The automated labelling of the hip abductor muscles, needed to estimate their FF, is based on a multi-atlas segmentation method, which was developed in C# and implemented in a plugin for Simpleware™ ScanIP (Version 2018.12; Synopsys, Inc., Mountain View, USA). In our implementation every atlas in the library is registered to the image to be segmented (target image); the registered atlases are sorted in descending similarity order; the labels of the 5 most similar atlases are propagated to the target image space; and subsequently fused into a single label for each muscle using majority voting (32). The atlas library consisted of 12 atlases, 10 of which corresponded to the HV and two were patients with relatively high BMI to add anatomical variability. Each atlas consisted of the in-phase Dixon image and a manually segmented labels image. The total number of atlases included in the library was chosen taking into account segmentation performance and processing time factors. In Appendix A, the impact of the library size on the segmentation performance is assessed.

Before running the multi-atlas segmentation, a data preparation stage is executed, where a bias field inhomogeneity correction filter (33) is applied to correct for low frequency intensity non-uniformities in the images.

For the image registration, a rigid followed by a *B-spline* non-rigid registration (34) was implemented using SimpleElastix (35, 36). The normalized cross-correlation (NCC) was used as similarity metric and the cost function (negative NCC) was minimized using the adaptive stochastic gradient descent algorithm (37) with 2000 iterations and 2048 samples. A pyramidal scheme of four layers with down-sampling factors of 8, 4, 2 and 1 was employed to improve the registration. These parameters have been previously optimized to achieve better segmentation performance.

The registered atlases were sorted in descending order of similarity to the target, which was quantified using the global normalized cross-correlation (GNCC). The 5 most similar atlases (with highest GNCC values) were selected and their labels propagated to be fused with majority voting. The number of selected atlases was chosen empirically based on segmentation performance (see Appendix A). Voxels where there was not a unique label with the highest number of votes were labelled as undecided, which were subsequently assigned to the closest label using distance maps.

In a post-processing stage, a soft-tissue intensity mask was applied to remove subcutaneous fat voxels from the labels. As IMF voxels could be excluded with this mask, a morphological close operation is applied to generate a soft-tissue mask that excludes background and subcutaneous fat tissue voxels, but preserve IMF voxels.

2.5 Evaluation

We assessed the performance of our method by comparing the FF obtained from the automatically segmented images to the manually segmented, for 17 individuals. In those cases where the target scan was one of the atlases in the library, the atlas was removed before executing the automated segmentation. The in-phase Dixon image was used as target image for every case.

We computed measures of average (mean and median) and spread (standard deviation (SD) and interquartile range (IQR)) for FF for each muscle for the manually and automatically segmented images for the full set of scans, and split by HV and patients groups. We generated boxplots for each of these groups.

A linear regression model was fit to the data to evaluate the correlation between the FF from the automated and manual segmentations, where the coefficient of determination (R^2) was used to indicate the level of correlation. In addition, we performed a Bland-Altman analysis that compared the FF from automated and manual segmentations for every muscle.

To quantify the overall accuracy of the FF measurements, we used the root mean square error (RMSE):

$$RMSE_m = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(FFA_{mi} - FFM_{mi})^2}$$

where $RMSE_m$ is the root mean square error in muscle *m* (between GMAX, GMED, GMIN and TFL), N is the total number of samples of muscles *m* analysed (equal to two times the number of scans analysed); and FFA_{mi} and FFM_{mi} are the frat fraction from automated and manual segmentations respectively for the sample *i* of muscle *m*.

In addition, the Dice Similarity Coefficient (DSC), the Relative Volume Difference (RVD), sensitivity and precision were used to assess the segmentation performance independently of the FF accuracy. Sensitivity and precision metrics were defined as:

$$sensitivity_l = \frac{TP}{TP + FN}$$

$$precision_l = \frac{TP}{TP + FP}$$

where TP, FP and FN are the number of true positive, false positive and false negative voxels respectively for each label *l*.

3. Results

In Figure 3, boxplots of the FF in each abductor muscle for the 17 cases assessed (34 muscles per boxplot) are shown for the manually and automatically segmented images. For the manually segmented images, the median (IQR) FF values were 19.6 (15.9-23.3)%, 13.8 (11.4-16.0)%, 14.9 (12.0-17.3)% and 14.7 (12.9-20.6)% for GMAX, GMED, GMIN and TFL muscles respectively; while we obtained median (IQR) of 20.0 (16.0-26.4)%, 14.3 (10.9-16.5)%, 15.5 (13.9-18.6)% and 16.2 (13.5-25.6)% for the automatically segmented images, showing good agreement between the two segmentations. The agreement was reasserted by the linear regression model fit to the data (Figure 4), where a strong linear correlation (R^2=0.93, $p < 0.001$, m=0.99) was found between the FF from automated and manual segmentations.

In Table 2, the median (IQR), mean (±SD) and RMSE values are reported for FF for both manual and automated segmentations for each groups of scans. The mean (±SD) DSC and RVD values, used to assess segmentation, are also presented in the table. In Figure 5, mean (±SD) values of the segmentation performance metrics are presented for each muscle, which show that the performance was variable between muscles, being more accurate in larger muscles as GMAX and GMED (mean DSC of 0.93 and 0.88 respectively), and less accurate and more variable in GMIN and TFL (mean DSC of 0.93 and 0.88 respectively). The segmentation Precision had a similar trend than the DSC values with mean values of 0.93, 0.89, 0.79 and 0.84 for each muscle respectively. This is a relevant metric for the FF measurement as represents how many voxels have been wrongly labelled as part of a given muscle. The relative lower segmentation performance for the smaller muscles had a big impact on the volume measurements (RVD values in Table 2). However, the FF RMSE, obtained from the same set of images, was relatively low for every muscle under study and with a lower variability compared to the DSC and Precision values. The FF RMSE for all subjects was 1.6%, 0.8%, 2.1% and 2.7% for GMAX, GMED, GMIN and TFL respectively.

A Bland-Altman analysis that compares the FF measurements for automated and manual segmentations is shown in Figure 6 for every muscle. The discrepancy was very low for GMED and low for GMAX with 95% confidence intervals of [-1.4%, 1.7%] and [3.1%, 3.4%] respectively; while higher for GMIN and TFL, which also had a mild positive bias of approximately 1%. Figure 6 also shows that a greater error was observed in the measurements of the patients group (plotted with circles), but for higher mean FF values.

Figure 7 shows boxplots for FF values for the HV and patients groups, for the manually (a) and automatically (b) segmented images. When analysing the results divided by groups, there was good agreement in the median (IQR) FF values for the HV and patients groups between the manually and automatically segmented images. For example, we automatically measured median (IQR) GMAX FF values of 19.5 (15.2-21.2)%, and 22.6 (16.4-28.2)% for the HV and patients groups respectively; while the reference values from the manually segmented images were 19.1 (15.5-21.2)% and 21.8 (16.9-28.4)%. The full results can be found in Table 2.

The good agreement between the automated and manual measurements can be also seen when comparing the FF values between the two groups under study, where the patients group had a higher FF in every muscle, especially in GMIN and TFL. The mean FF difference between the patients and HV groups in the automated measurements was 2.9%, 3.7%, 5.5% and 6% for GMAX, GMED, GMIN and TFL respectively; while 3.7%, 4.0%, 8.0% and 4.3% for the manually segmented images. These FF differences between the two groups means a higher fat infiltration in the patients group in the range of 20-50% relatively to the HV group. In Figure 8, we show two cases for each group with labels of the FF from manual and automated segmentations next to each muscle. Case A is a good representation of the average subject in the HV group, while subject C corresponds to the outlier measurements in Figure 7 for GMAX and TFL in the HV group. For the patients group, we chose cases without (B) and with high fat infiltration (D). The images also show the outline of the labels from the automated segmentation.

4. Discussion

This is, to the best of our knowledge, the first work to evaluate the automated measurement of FF of the main hip abductor muscles using a multi-atlas based method and Dixon MRI. The proposed method delivered similar values to a manual method in terms of mean (±SD) FF values for each muscle and RMSE. The FF measurements also showed good agreement with those obtained from manually segmented images in terms of correlation and values distribution in the HV and patients group. The use of a multi-atlas segmentation method to label each muscle and Dixon imaging proved to be accurate enough to assess muscle composition, despite having a sub-optimal accuracy in terms of muscle volume.

4.1 Fat Fraction Accuracy

In terms of accuracy, the FF RMSE, when using the FF from manually segmented images as a reference, was low for the four abductors muscles, although less accurate in GMIN and TFL. Similarly, the segmentation performance was lower for GMIN and TFL, but in this case the difference was higher respect to GMAX and GMED. The lower segmentation performance in these two muscles had a relative low impact on the FF estimation but a high impact on the volume measurements (high RVD values). The lower impact of the segmentation errors on the FF estimation was expected as a wrongly labelled voxel introduces only a partial error since the FF for a given muscle is a mean value within a set of voxels, which have a reduced range of values (approximately 10%-30% in most of the cases). Another reason for the higher FF accuracy compared to the volume measurements is that the majority voting label fusion strategy and the post-processing mask tends to reduce the sensitivity and increase the precision of the segmentation, which is beneficial for FF estimation but produces an underestimation of the computed volume.

4.2 Segmentation Accuracy

When comparing our segmentation method to the small number of works available in the literature on hip abductor muscles segmentation, we obtained similar or marginally better performance in terms of DSC values. For example, Ranzini et al (38) used a combination of MRI and CT images to segment the hip muscles in patients with total hip replacement, where mean DSC values of 0.91, 0.85, 0.83 and 0.80 were obtained for GMAX, GMED, GMIN and TFL on the healthy side; but lower DSC values were reported when using only MR images. In the same context, Yokota et al (39) presented a multi-atlas segmentation method for the segmentation of the hip and thigh muscles from CT images, achieving mean DSC values of 0.89, 0.82 and 0.64 for GMAX, GMED and GMIN respectively, and 0.92, 0.87 and 0.70 when using a computationally demanding multi-stage method. Baudin et al (40) reported a median DSC values of 0.80 for the segmentation of TFL from MRI but with high variability, including cases with very low DSC values (from 0.1 to 0.5). IMF was measured only in the first of these works, however the authors used an intensity-based method from T1-weighted images, which is not quantitative as our FF measurements from Dixon images, and therefore its accuracy cannot be quantified.

4.3 Differences between Healthy Volunteers and Patients

When comparing the HV with the patients group, we observed a considerable increase of the FF in the four abductor muscles for the patients group for both manually and automatically segmented images. This is concordant with data on FF measurement of thigh muscles, which has shown a correlation between FF and muscle health (3, 6, 41). In this preliminary study, we present and evaluate a method that can accurately measure FF in the abductor muscles to compare two different groups of individuals. The increased FF levels in the patient group could be due to a reduced level of mobility, but also because of age differences between the two groups, as fat infiltration in the thigh and calves has been associated with sarcopenia and aging (3, 6, 31).

4.4 Differences between Muscles

In the HV group, the fat content in GMAX was higher than in the other abductor muscles and this was observed equally for both the automated and manual segmentations. The higher IMF in GMAX can be also noticed by visual inspection in Figure 8. GMED and GMIN had similar FF values, which is not surprising as they are functionally equivalent and have similar characteristics. On the other hand, GMAX, which is a powerful extensor of the hip, has a different functionality and fibre composition than the other gluteal muscles (42), and this could explain the difference in FF values. Differences in IMF content within a muscle group have also been detected in the calves (29, 31, 43).

4.5 Fat Fraction vs Volume

FF proved to be a suitable metric for automated and quantitative assessment of muscles. Another potential metric to automatically evaluate individual muscles is volume (44, 45), however the measurement of volume requires a higher accuracy as changes in muscle size are much smaller between healthy subjects and patients (46, 47) than for FF. An additional advantage of FF is that it is independent of the subject size and hence suitable for establishing baseline values for healthy subjects.

4.6 Limitations

A limitation of the present method is the small number of patients included in the atlas library, which had impact on the FF accuracy for this group. Increasing the number of subjects in the atlas library could also overcome the lower accuracy in the TFL muscle by accounting for the greater anatomical variability of this muscle.

A second limitation of this work is the inclusion of only patients with OA. Patients with more severe disease would present a higher muscle fat infiltration that could involve a greater proportion of the full muscle. This would present a greater challenge for the multi-atlas segmentation algorithm and demand the introduction of new strategies to address their automatic segmentation.

5. Conclusion

We present a multi-atlas based method that automatically estimates FF in the hip abductor muscles from Dixon MR images, as a measure of fat infiltration. The method showed very good accuracy and agreement with the FF from manually segmented images. The error in the FF measurements was low. The mean FF in the hip abductors was considerable higher in a small group of orthopaedic patients than in healthy volunteers. This solution adds a further tool to enable clinicians, physiotherapists and sport scientists to measure and monitor the results of their various surgical and exercise interventions aimed at the rehabilitation patients with musculoskeletal disease.

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgments

This research study was funded by The Maurice Hatter Foundation, the RNOH Charity, the Rosetrees Trust and the Stoneygate Trust and supported by researchers at the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Bibliography

1. Anderson FC, Pandy MG: Individual muscle contributions to support in normal walking. *Gait Posture* 2003; 17:159–169.

2. Bartlett JL, Sumner B, Ellis RG, Kram R: Activity and functions of the human gluteal muscles in walking, running, sprinting, and climbing. *Am J Phys Anthropol* 2014; 153:124–131.

3. Addison O, Marcus RL, Lastayo PC, Ryan AS: Intermuscular fat: A review of the consequences and causes. *Int J Endocrinol* 2014.

4. Cruz-Jentoft AJ, Bahat G, Bauer J, et al.: Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing* 2019; 48:16–31.

5. Marcus RL, Addison O, Kidde JP, Dibble LE, Lastayo PC: Skeletal muscle fat infiltration: Impact of age, inactivity, and exercise. *J Nutr Heal Aging* 2010; 14:362–366.

6. Marcus RL, Addison O, Dibble LE, Foreman KB, Morrell G, Lastayo P: Intramuscular adipose tissue, sarcopenia, and mobility function in older individuals. *J Aging Res* 2012; 2012.

7. Heffler MA, Barlow R, Xi Y, et al.: Multi-parametric muscle and fat correlation of

computed tomography parameters to outcomes in a total hip arthroplasty population. *BMC Musculoskelet Disord* 2018; 19:4.

8. Kiyoshige Y, Watanabe E: Fatty degeneration of gluteus minimus muscle as a predictor of falls. *Arch Gerontol Geriatr* 2015; 60:59–61.

9. Emery AE: The muscular dystrophies. *Lancet* 2002; 359:687–695.

10. Lareau-Trudel E, Troter A Le, Ghattas B, et al.: Muscle quantitative MR imaging and clustering analysis in patients with facioscapulohumeral muscular dystrophy type 1. *PLoS One* 2015; 10.

11. Wokke BH, Bos C, Reijnierse M, et al.: Comparison of dixon and T1-weighted MR methods to assess the degree of fat infiltration in duchenne muscular dystrophy patients. *J Magn Reson Imaging* 2013; 38:619–624.

12. Figueroa-Bonaparte S, Llauger J, Segovia S, et al.: Quantitative muscle MRI to follow up late onset Pompe patients: A prospective study. *Sci Rep* 2018; 8.

13. 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.

14. Bley TA, Wieben O, François CJ, Brittain JH, Reeder SB: Fat and water magnetic resonance imaging. *J Magn Reson Imaging* 2010; 31:4–18.

15. Xiang QS, An L: Water-fat imaging with direct phase encoding. *J Magn Reson Imaging* ; 7:1002–15.

16. Hernando D, Liang Z-P, Kellman P: Chemical shift-based water/fat separation: A comparison of signal models. *Magn Reson Med* 2010; 64:811–822.

17. Dixon WT: Simple proton spectroscopic imaging. *Radiology* 1984; 153:189–194.

18. Kovanlikaya A, Mittelman SD, Ward A, Geffner ME, Dorey F, Gilsanz V: Obesity and fat quantification in lean tissues using three-point Dixon MR imaging. *Pediatr Radiol* 2005; 35:601–607.

19. 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. *J Magn Reson Imaging* 2015; 41:1558–1569.

20. Le Troter A, Fouré A, Guye M, et al.: Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches. *Magn Reson Mater Phy* 2016; 29:245–257.

21. Scheys L, Loeckx D, Spaepen A, Suetens P, Jonkers I: Atlas-based non-rigid image registration to automatically define line-of-action muscle models: A validation study. *J Biomech* 2009; 42:565–572.

22. Ghosh S, Ray N, Boulanger P: A Structured Deep-Learning Based Approach for the Automated Segmentation of Human Leg Muscle from 3D MRI. In *2017 14th Conf Comput Robot Vis*. IEEE; 2017:117–123.

23. 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 *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. IEEE; 2017:317–320.

24. Kolk S, Klawer EME, Schepers J, Weerdesteyn V, Visser EP, Verdonschot N: Muscle Activity during Walking Measured Using 3D MRI Segmentations and [18F]-Fluorodeoxyglucose in Combination with Positron Emission Tomography. *Med Sci Sports Exerc* 2015; 47:1896–1905.

25. Leijendekkers RA, Marra MA, Ploegmakers MJM, et al.: Magnetic-resonance-imaging-based three-dimensional muscle reconstruction of hip abductor muscle volume in a person with a transfemoral bone-anchored prosthesis: A feasibility study. *Physiother Theory Pract* 2019; 35:495–504.

26. Ranzini MBM, Henckel J, Ebner M, et al.: Automated postoperative muscle assessment of hip arthroplasty patients using multimodal imaging joint segmentation. *Comput Methods Programs Biomed* 2020; 183:105062.

27. Yokota F, Otake Y, Takao M, et al.: Automated muscle segmentation from CT images of the hip and thigh using a hierarchical multi-atlas method. *Int J Comput Assist Radiol Surg* 2018; 13:977–986.

28. Goodpaster BH, Park SW, Harris TB, et al.: The loss of skeletal muscle strength, mass, and quality in older adults: The Health, Aging and Body Composition Study. *Journals Gerontol - Ser A Biol Sci Med Sci* 2006; 61:1059–1064.

29. Hilton TN, Tuttle LJ, Bohnert KL, Mueller MJ, Sinacore DR: Excessive Adipose Tissue Infiltration in Skeletal Muscle in Individuals With Obesity, Diabetes Mellitus, and Peripheral Neuropathy: Association With Performance and Function. *Phys Ther* 2008; 88:1336–1344.

30. Manini TM, Clark BC, Nalls MA, Goodpaster BH, Ploutz-Snyder LL, Harris TB: Reduced physical activity increases intermuscular adipose tissue in healthy young adults. *Am J Clin Nutr* 2007; 85:377–384.

31. Tuttle LJ, Sinacore DR, Mueller MJ: Intermuscular Adipose Tissue Is Muscle Specific and Associated with Poor Functional Performance. *J Aging Res* 2012; 2012.

32. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage* 2009;

46:726–738.

33. Tustison NJ, Avants BB, Cook PA, et al.: N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging* 2010; 29:1310–1320.

34. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ: Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 1999; 18:712–721.

35. Marstal K, Berendsen F, Staring M, Klein S: SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration. In *2016 IEEE Conf Comput Vis Pattern Recognit Work*. IEEE; 2016:574–582.

36. Klein S, Staring M, Murphy K, Viergever MA, Pluim J: elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans Med Imaging* 2010; 29:196–205.

37. Vandervoort E, Sossi V: An analytical scatter correction for singles-mode transmission data in PET. *IEEE Trans Med Imaging* 2008; 27:402–412.

38. Ranzini MBM, Henckel J, Ebner M, et al.: Automated postoperative muscle assessment of hip arthroplasty patients using multimodal imaging joint segmentation. *Comput Methods Programs Biomed* 2020; 183:105062.

39. Yokota F, Otake Y, Takao M, et al.: Automated muscle segmentation from CT images of the hip and thigh using a hierarchical multi-atlas method. *Int J Comput Assist Radiol Surg* 2018; 13:977–986.

40. Baudin P-Y, Azzabou N, Carlier PG, Paragios N: Automatic skeletal muscle segmentation through random walks and graph-based seed placement. In *2012 9th IEEE Int Symp Biomed Imaging*. IEEE; 2012:1036–1039.

41. Yoshida Y, Marcus RL, Lastayo PC: Intramuscular adipose tissue and central activation

in older adults. *Muscle Nerve* 2012; 46:813–816.

42. Johnson MA, Polgar J, Weightman D, Appleton D: Data on the distribution of fibre types in thirty-six human muscles. An autopsy study. *J Neurol Sci* 1973; 18:111–129.

43. Commean PK, Tuttle LJ, Hastings MK, Strube MJ, Mueller MJ: Magnetic resonance imaging measurement reproducibility for calf muscle and adipose tissue volume. *J Magn Reson Imaging* 2011; 34:1285–94.

44. O'Brien TD, Reeves ND, Baltzopoulos V, Jones DA, Maganaris CN: Strong relationships exist between muscle volume, joint power and whole-body external mechanical power in adults and children. *Exp Physiol* 2009; 94:731–738.

45. Belavý DL, Miokovic T, Rittweger J, Felsenberg D: Estimation of changes in volume of individual lower-limb muscles using magnetic resonance imaging (during bed-rest). *Physiol Meas* 2011; 32:35–50.

46. Smith MMF, Bonacci J, Mendis MD, Christie C, Rotstein A, Hides JA: Gluteus medius activation during running is a risk factor for season hamstring injuries in elite footballers. *J Sci Med Sport* 2017; 20:159–163.

47. Zacharias A, Pizzari T, English DJ, Kapakoulakis T, Green RA: Hip abductor muscle volume in hip osteoarthritis and matched controls. *Osteoarthr Cartil* 2016; 24:1727–1735.

Appendix A – Library Size and Atlas Selection

The library size and number of selected atlases for label fusion and propagation was optimized by comparing mean (±SD) DSC values for different configurations. Library sizes from 8 to 17 atlases were evaluated, where for the case of size 17, the effective size of the library was 16 as each case to be segmented was removed from the library before running the segmentation. When incrementing the library size, first the 10 HV scans were used, followed by the patients with highest BMI (as the HV had a lower BMI) and keeping gender approximately balanced

For each library size, different number of selected atlases for label propagation and fusion were compared. In Figure A.1, the mean (±SD) DSC values for all the muscle labels are plot for different library sizes and number of selected atlases. We selected a library size of 12 atlases and the use of the 5 most similar atlases in the label fusion as the optimal configuration since there were only marginal gains when increasing the library size and it came at the cost of higher segmentation times (computation times are approximately proportional to the number of atlases in the library).

Tables

	Volunteers	*N*	*Age [years]*
Group 1 Healthy Volunteers	Female	6	31.2 (20-43)
	Male	4	27.0 (22-35)
Group 2 Patients	Female	4	60.5 (37-77)
	Male	3	64.0 (45-75)

Table 1. Demographics of the healthy volunteers and patients group. The age values correspond to mean (min-max) values.

			GMAX		GMED		GMIN		TFL	
			Manual	Automated	Manual	Automated	Manual	Automated	Manual	Automated
Healthy Volunteers	Fat Fraction [%]	Mean (±SD)	19.9 (6.0)	19.7 (5.5)	12.9 (3.0)	12.9 (3.0)	14.0 (2.8)	14.1 (2.8)	15.8 (6.8)	17.5 (7.7)
		Median (IQR)	19.1 (15.5-21.2)	19.5 (15.2-21.2)	11.9 (10.6-15.6)	13.0 (10.5-15.6)	14.3 (11.8-16.3)	14.7 (11.4-16.0)	14.2 (12.2-16.5)	15.7 (13.7-19.5)
		RMSE [%]	0	1.3	0	0.7	0	0.8	0	2.3
	DSC	Mean (±SD)	1	0.94 (0.02)	1	0.88 (0.03)	1	0.83 (0.04)	1	0.81 (0.05)
	RVD	Mean (±SD)	0	-1.9 (3.8)	0	-1.8 (8.8)	0	2.7 (12.3)	0	-3.8 (19.5)
Patients	Fat Fraction [%]	Mean (±SD)	22.8 (6.4)	23.4 (6.7)	16.6 (4.5)	16.9 (4.3)	19.5 (7.2)	22.1 (8.1)	21.8 (9.0)	21.8 (8.4)
		Median (IQR)	21.8 (16.9-28.4)	22.6 (16.4-28.2)	15.1 (12.0-20.8)	15.4 (13.2-21.3)	15.4 (14.4-26.9)	19.4 (15.0-29.9)	19.7 (13.5-27.9)	24.4 (12.6-27.8)
		RMSE [%]	0	2.1	0	0.9	0	2.3	0	3.1
	DSC	Mean (±SD)	1	0.91 (0.01)	1	0.88 (0.02)	1	0.78 (0.05)	1	0.76 (0.06)
	RVD	Mean (±SD)	0	-2.1 (6.5)	0	-5.2 (6.7)	0	3.8 (15.8)	0	-16.4 (21.1)
Healthy Volunteers + Patients	Fat Fraction [%]	Mean (±SD)	21.1 (6.3)	21.2 (6.2)	14.4 (4.1)	14.6 (4.1)	16.2 (5.7)	17.4 (6.8)	18.3 (8.2)	19.3 (8.1)
		Median (IQR)	19.6 (15.9-23.3)	20.0 (16.0-26.4)	13.8 (11.4-16.0)	14.3 (10.9-16.5)	14.9 (12.0-17.3)	15.5 (13.9-18.6)	14.7 (12.9-20.6)	16.2 (13.5-25.6)
		RMSE [%]	0	1.6	0	0.8	0	2.1	0	2.7
	DSC	Mean (±SD)	1	0.93 (0.02)	1	0.88 (0.02)	1	0.81 (0.05)	1	0.79 (0.06)
	Sensitivity	Mean (±SD)	1	0.93 (0.03)	1	0.87 (0.05)	1	0.83 (0.06)	1	0.78 (0.12)
	Precision	Mean (±SD)	1	0.93 (0.04)	1	0.89 (0.02)	1	0.79 (0.09)	1	0.84 (0.08)
	RVD	Mean (±SD)	0	-2.0 (5.0)	0.0	-3.1 (8.1)	0.0	3.1 (13.6)	0.0	-9.0 (20.9)

Table 2. Overall results for fat fraction from manually and automatically segmented images. The mean (±SD), median (IQR) and RMSE values are reported for healthy volunteers, patients and all subjects together. The DSC, Sensitivity, Precision and RVD values for the segmentation performance are also reported.

Figures

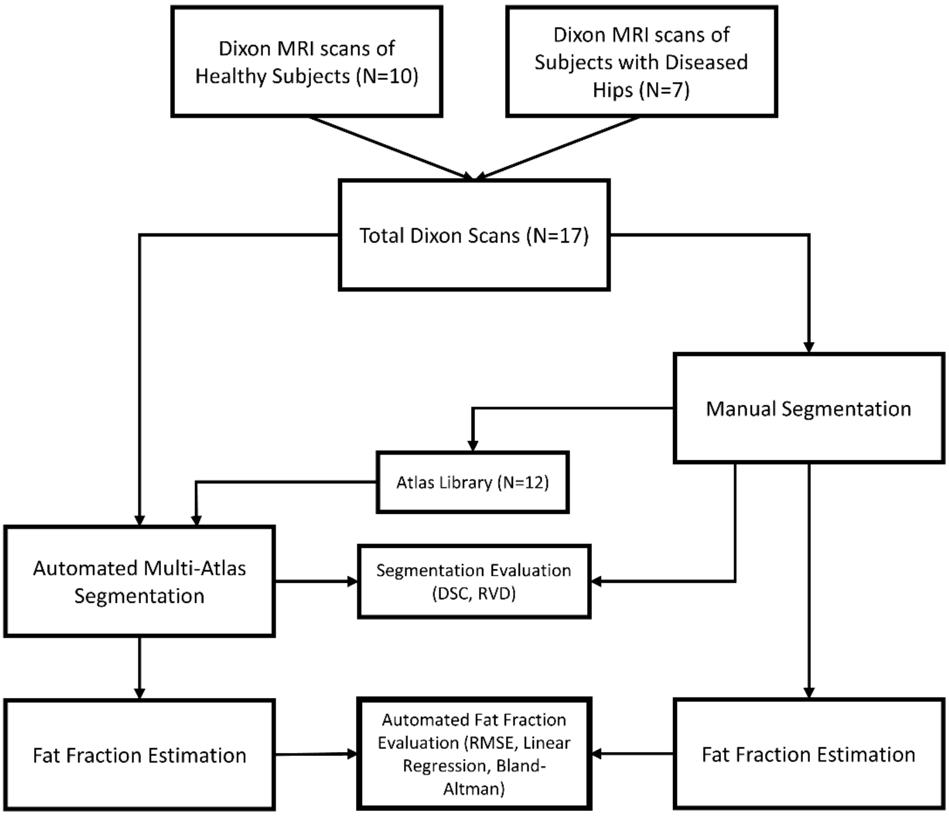

Figure 1. Flowchart describing the procedure to implement and evaluate our method for automated measurement of fat fraction in the hip abductor muscles.

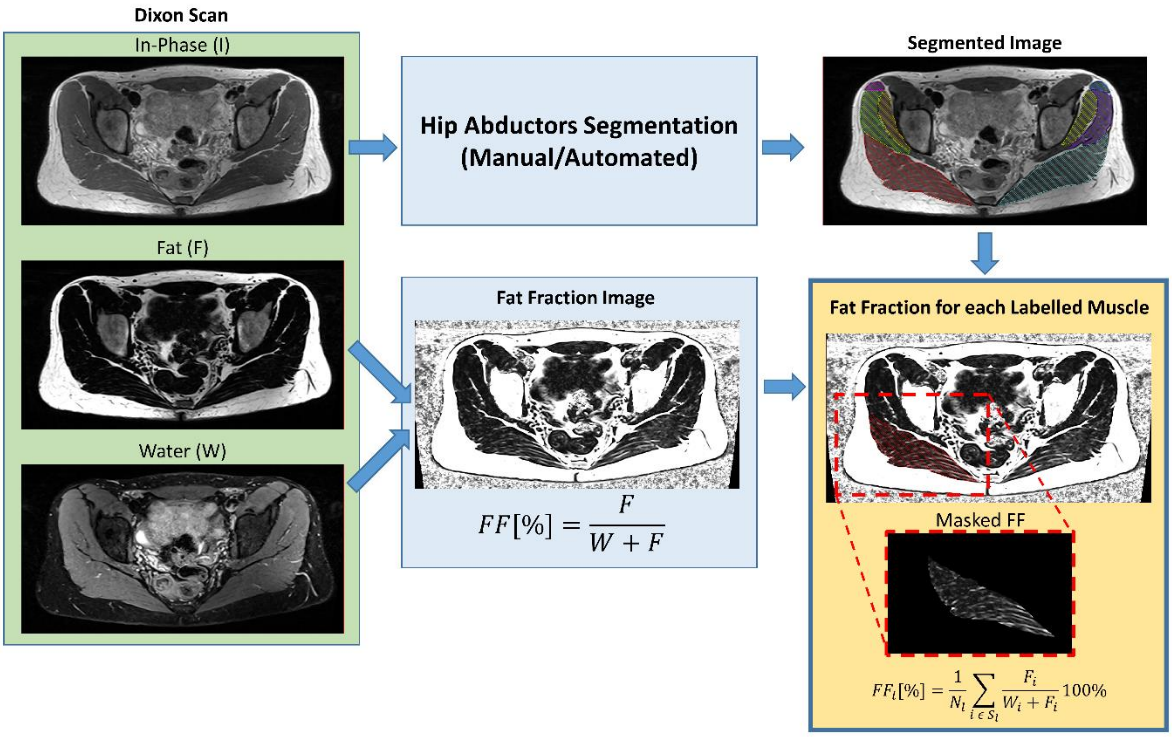

Figure 2. Estimation of fat fraction in the hip abductors muscles from a Dixon scan. The in-phase image is used in the segmentation of the hip abductors, which can be done manually or automatically. A FF image is obtained from the fat and water Dixon images. For each segmented muscle, its label is applied as a mask to the FF image and the mean FF is estimated within the mask voxels.

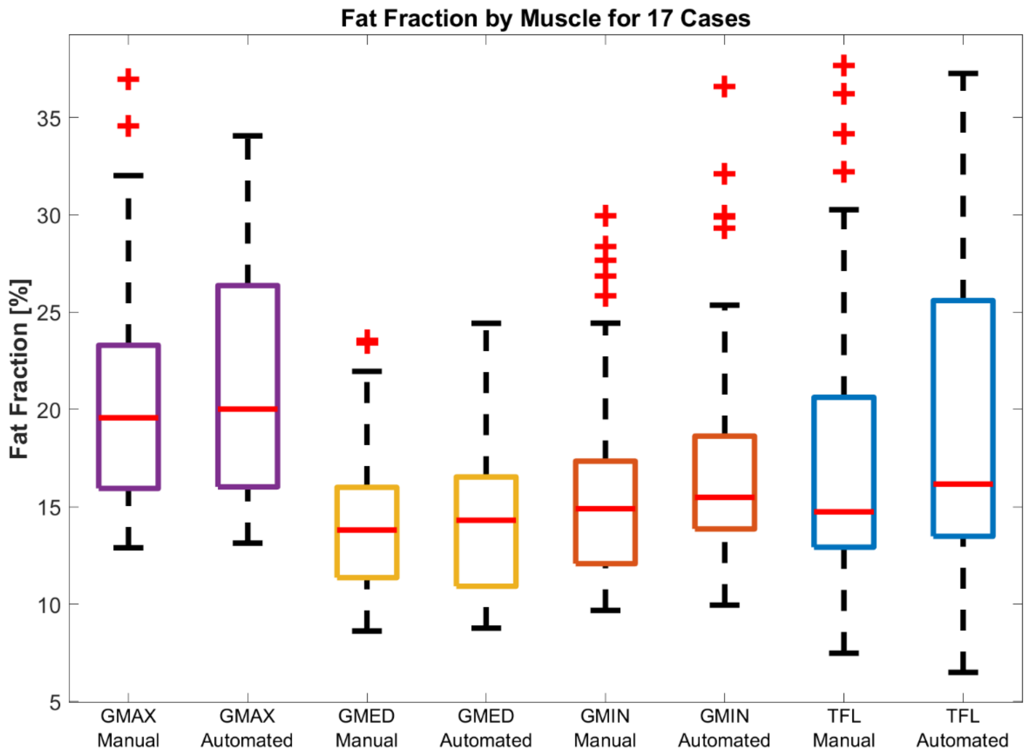

Figure 3. Boxplots of the FF in each of the hip abductor muscles for 17 cases (34 muscles). The FF boxplots from the manually and automatically segmented are shown next to each other for each muscle.

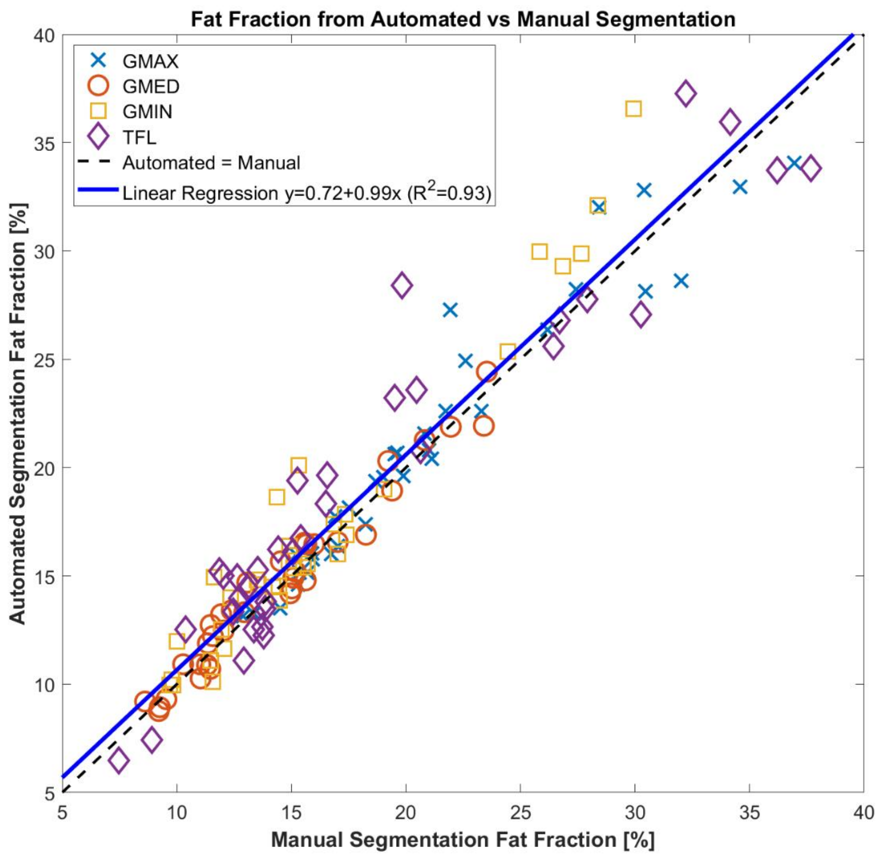

Figure 4. FF in each of the hip abductor muscles from automated segmentations plotted against FF from manually segmented images. A different marker is used for each muscle. The case for automated=manual is shown in a dashed line and a liner regression fit to the data in a solid line.

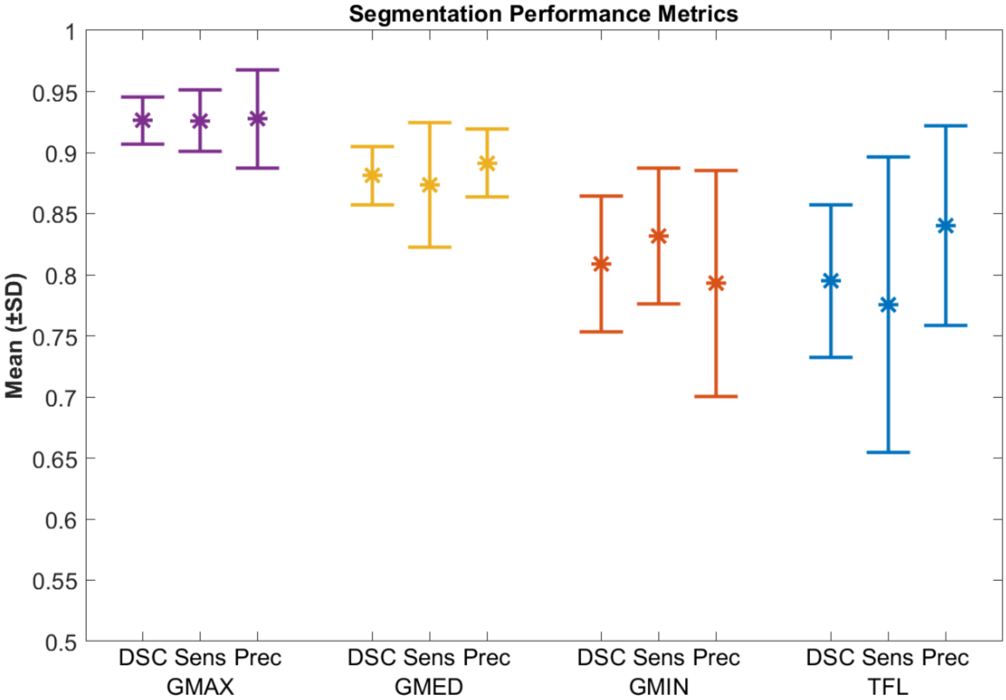

Figure 5. Mean (±SD) of DSC, Sensitivity and Precision segmentation performance metrics for GMAX, GMED, GMIN and TFL.

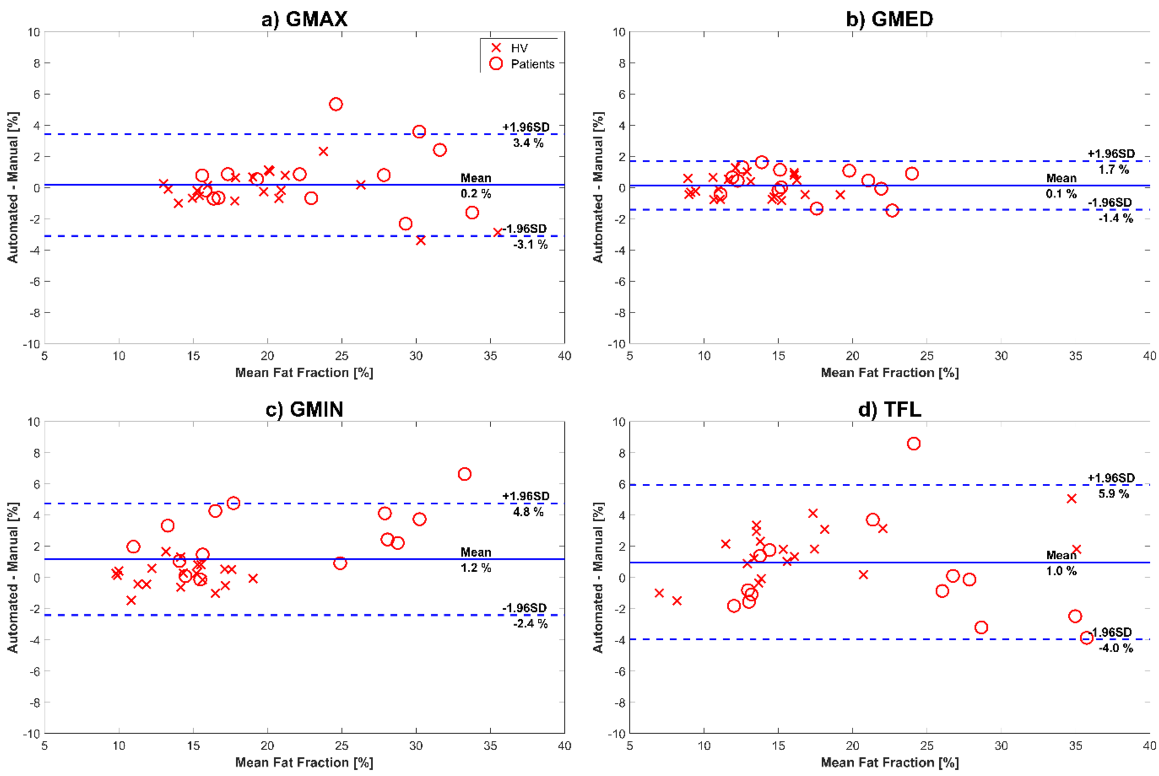

Figure 6. Bland-Altman analysis of the FF for each muscle comparing automated and manual segmentations. On the *x* axis the mean FF for each case and in the *y* axis the FF discrepancy between automated and manual segmentations. The HV cases are with crosses, while the patients with circles.

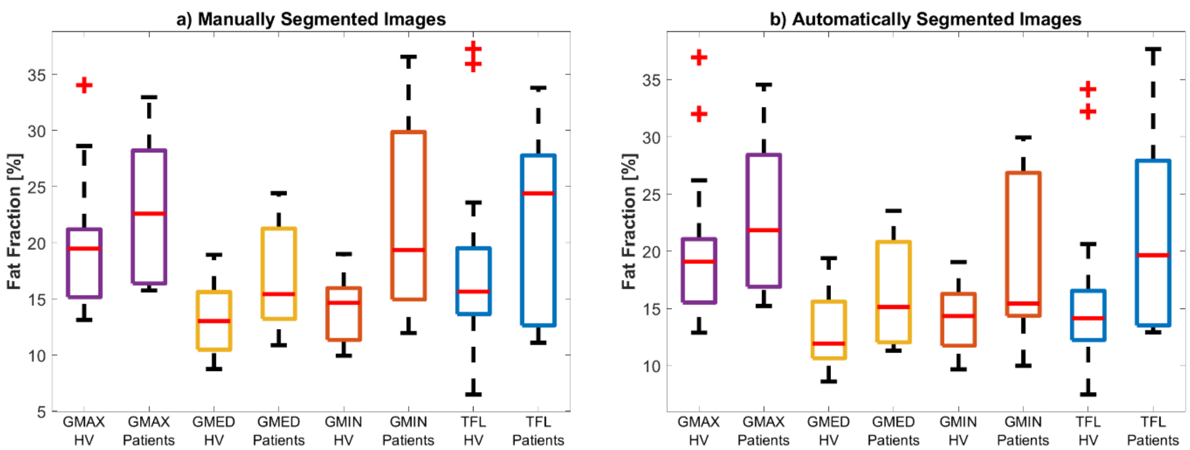

Figure 7. Boxplots of FF in each muscle for the HV and patients groups from a) manually and b) automatically segmented images.

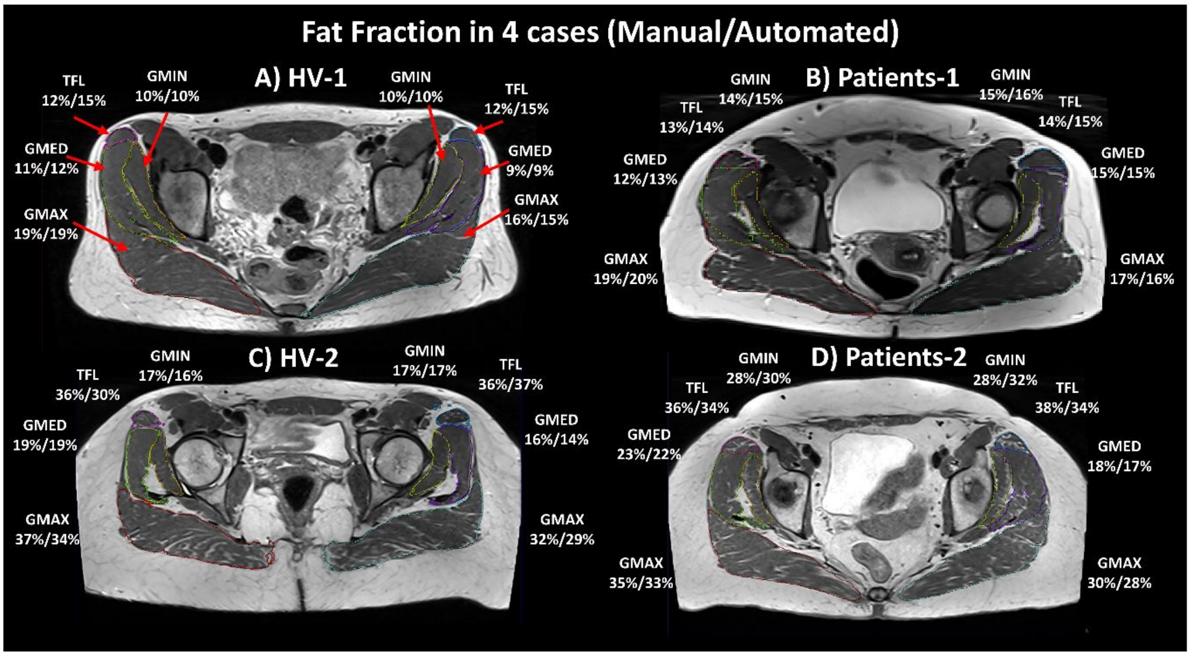

Figure 8. Example of FF values in 2 cases from the HV group (A and C) and 2 from the patients group (B and D). The FF values from the manual and automatically segmented images are shown for every muscle (Manual/Automated). The labels for each muscle correspond to the automated segmentation.

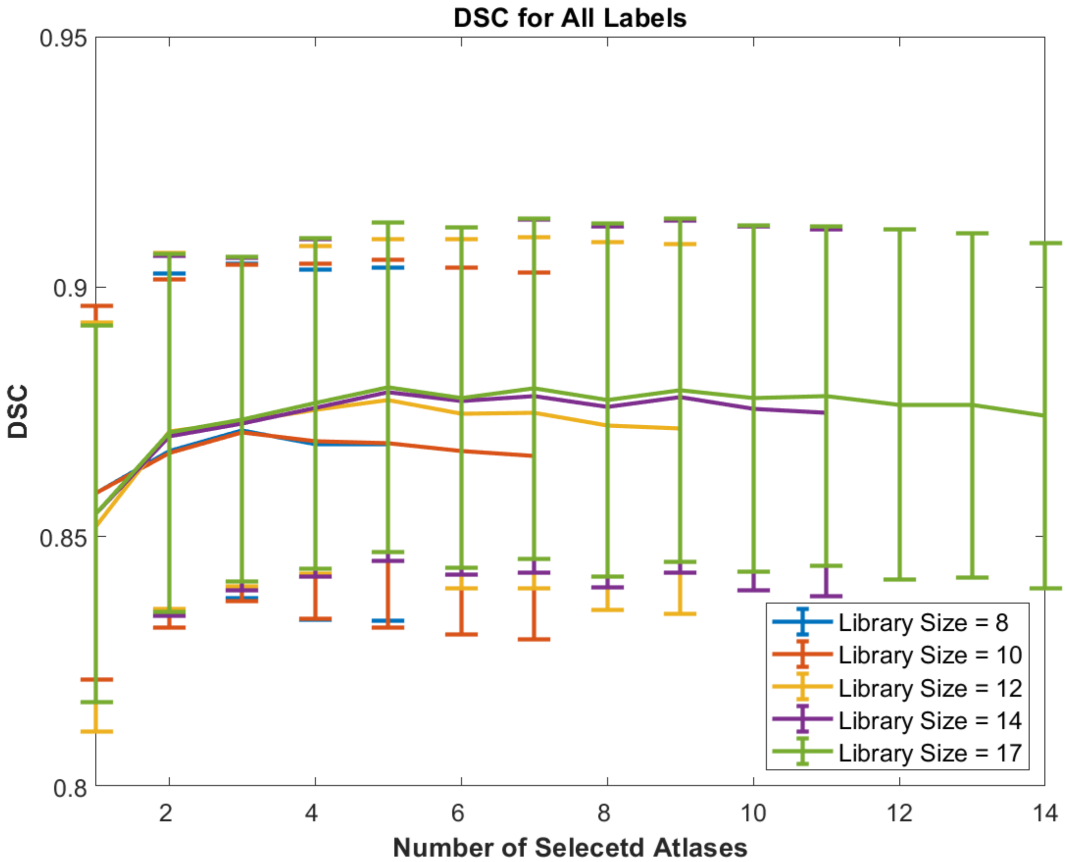

Figure A.1. Mean (±SD) DSC values as a function of the number of selected labels for label fusion, for different library sizes.