Test–retest reliability of rapid whole body and compartmental fat volume quantification on a widebore 3T MR system in normal-weight, overweight, and obese subjects


Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:
http://dx.doi.org/10.1002/jmri.25326
Copyright: Wiley
http://eu.wiley.com/WileyCDA/

Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-128881
Title

Test-retest reliability of rapid whole body and compartmental fat volume quantification on a wide bore 3T MR system in normal-weight, overweight and obese subjects.

Authors

David Newman FRCR¹
Christian Kelly-Morland FRCR¹
Olof Dahlqvist Leinhard PhD²,³,⁵
Bahman Kasmai MSc¹
Richard Greenwood MSc¹
Paul N Malcolm FRCR¹
Thobias Romu MSc²,⁴,⁵
Magnus Borga PhD²,⁴,⁵
Andoni P Toms FRCR PhD¹

Affiliations

1 Department of Radiology
   Norfolk & Norwich University Hospital, United Kingdom

2 Center for Medical Image Science and Visualisation
   Linköping University, Sweden

3 Department of Medical and Health Sciences
   Linköping University, Sweden

4 Department of Biomedical Engineering
   Linköping University, Sweden

5 Advanced MR Analytics AB
   Badhusgatan 5, Linköping, Sweden

Correspondence

Dr Andoni P Toms, Consultant Radiologist
Radiology Academy
Cotman Centre
Colney Lane
Norwich, Norfolk NR4 7UB
United Kingdom
Telephone 44 1603 286143
Fax 44 1603 286077
Email andoni.toms@nnuh.nhs.uk
Running Title

Reliability of MR fat quantification
ABSTRACT

Purpose

To measure the test-retest reliability of rapid (< 15 min) whole body and visceral fat volume quantification in normal and obese subjects on a wide bore 3T MR system and compare with conventional manual segmentation.

Materials and Methods

Thirty participants (BMI 20.1–48.6 kg/m²) underwent two whole body MRI examinations on a wide bore 3T machine using a 2-point Dixon technique. Phase sensitive reconstruction and intensity inhomogeneity correction produced quantitative datasets of Total Adipose Tissue (TAT), Abdominal Subcutaneous Adipose Tissue (ASAT) and Visceral Adipose Tissue (VAT). The quantification was performed automatically using non-rigid atlas-based segmentation and compared with manual segmentation (SliceOmatic™).

Results

The mean TAT was 31.74 L with a coefficient of variation (CV) of 0.79% and a coefficient of repeatability (CR) of 0.49 L. The ASAT was 7.92 L with a CV of 2.98% and a CR of 0.46 L. There was no significant difference in the semi-automated and manually segmented VAT (p=0.73) but there were differences in the reliability of the two techniques. The mean semi-automated VAT was 2.56 L, CV 1.8% and CR 0.09 L compared to the mean manually segmented VAT of 3.12 L where the CV was 6.3% and the CR was 0.39 L.

Conclusion

Rapid semi-automated whole body and compartmental fat volume quantification can be derived from a wide bore 3T system, for a range of body sizes including obese patients, with “almost perfect” test-retest reliability.
KEYWORDS

- Semi-automated quantification
- Manual segmentation
- MRI
- Adipose
- Visceral adipose
- Reliability
INTRODUCTION

Obesity has reached epidemic proportions with about 1.5 billion adults classified as overweight in 2008 (1). Excess body weight is associated with increased risks including type 2 diabetes, hypertension, neurovascular and cardiovascular diseases (2–4). These disorders often result in an increased risk of early death and significant reduction in quality of life. As well as the risks associated with total body weight, in effect total body fat, there are risks associated with particular patterns of fat distribution. Visceral adipose tissue (VAT), or fat distributed in and around organs, appears to be more important than non-visceral fat in the development of insulin resistance (5–8) and cardio-vascular disease (9). Measuring the amount of fat within different compartments in the body can only be done with several imaging techniques. While this can be accomplished with CT and, to some extent, dual energy x-ray absorptiometry (DXA) (10, 11), the major focus of recent research has been in developing robust MR imaging techniques that obviate the need for exposure to ionizing radiation.

Methods for quantifying body fat have been successfully developed over the past 20 years (12). Measurement of gross body fat in separate compartments was originally very time-consuming because of the need to manually segment the different compartments. More recently introduced semi-automated and automated segmentation techniques have made this process faster (13–17) and there is increasing evidence of good reliability and correlation with the laborious standard of manual segmentation (18). Variances of up to 1% for semi-automated segmentation of superficial fat, and 1–2% for intramuscular adipose tissue (19) have been reported and test-retest reliability appears to be excellent with a study of visceral and subcutaneous adipose tissue in children using a fully automated process demonstrating correlation coefficients of r=0.977 for VAT and r=0.999 for subcutaneous adipose tissue (SAT)(20).

Wide bore MR systems are being installed more frequently, partly to address the imaging of larger patients and research participants. Wide bore machines pose significant technical challenges in producing a homogeneous static magnetic field, particularly at the peripheries, which is further complicated by unpredictable field distortions produced by the patient. While the reliability of semi-automated fat quantification has been established in normal and overweight patients (21–23) the reliability in obese patients in a widebore MR system has not yet been assessed.

The primary aim of this study is to measure the test-retest reliability of a new semi-automated MR protocol designed to measure whole body adipose tissue, abdominal subcutaneous adipose tissue (ASAT) and visceral adipose tissue (VAT) content on a wide bore 3T MR system. The secondary aim was to perform method comparison of semi-automated VAT quantification with an approved industry standard technique.
MATERIALS AND METHODS

This is the second part of a prospective test-retest reliability and method comparison study. Ethical approval for this study was obtained from the local Research Ethics Review Committee on 29th November 2012. Informed written consent was obtained from each participant. The first part of the study investigated the performance of semi-automated quantification of muscle (24) whereas this part investigates the performance of fat quantification.

The required sample size of a minimum 19 volunteers was calculated assuming a minimal acceptable test-retest correlation coefficient $\rho = 0.7$, $\alpha = 0.05$, $\beta = 0.2$, then for a final test-retest correlation coefficient $\rho = 0.9$ using two repeated measures (25). The final sample size of n=30 was chosen to achieve this and to allow for a spread of body mass indices (BMI) from normal to obese in both genders.

15 female and 15 male were recruited from Radiology department staff at the Norfolk & Norwich University Hospital between December 2012 and February 2013. Five participants from each gender were voluntarily recruited for each of the following categories of BMI; 18–25 kg/m$^2$ (normal), 25–30 kg/m$^2$ (overweight) and >30 kg/m$^2$ (obese). The volunteers’ age, gender, height, weight and scan times were recorded on the day of the MR examinations.

MRI protocol

Each participant underwent whole body fat and water MR imaging using a modified 2–point Dixon fat and water sequence on a MR750w wide bore 3T MR machine (GE Medical Systems Ltd., Hatfield, UK) using the integrated quadrature body coil. MR images were acquired with the volunteers supine and with arms by their sides. The pulse sequence was a 3D dual echo spoiled gradient echo sequence with in and opposed phase echo times of 1.15 ms and 2.3 ms respectively. The repetition time was 3.7ms and the flip angle was 10° with an acquired resolution of 2.5 x 2.8 x 3 mm$^3$ and a field of view of 50 x 50 cm. The image protocol was applied repeatedly; starting from the head, with a 30 mm image stack overlap until whole body coverage was achieved using nine separate stations. Five stations covered the thorax, abdomen and pelvis each acquired with a 20 s breathhold. The head, neck and lower limbs were acquired with four non-breathhold acquisitions of between 27 s and 34 s scan time.

The total examination time was 10–15 minutes. After the first scan each participant was removed from the MR machine, moved to a changing facility, and after an interval of at least 20 minutes then re-imaged.
Post-processing

Anonymized DICOM datasets were processed at the Centre for Medical Image Science and Visualization (CMIV) at Linköping University. Semi-automated segmentation was performed at CMIV and manual segmentation was performed in Norwich.

Semi-automated quantification technique

The semi-automated image analysis was performed using software provided by AMRA™ (Advanced MR Analytics AB, Linköping, Sweden); this method has been previously described (24, 26–28) and is used in AMRA Profiler™, a cloud-based analysis service for MR-based body composition analysis. In brief, the analysis consists of the following steps:

Phase sensitive reconstruction was performed to remove variable phase offset resulting in paired pure fat and water based images (29). These were calibrated by calculating quantitative fat images based on pure adipose tissue as an internal intensity reference. As a result the signal intensity level in a given fat image voxel is related to the intensity in pure adipose tissue and is given the value 1 corresponding to 100% adipose tissue (26). In this way adipose tissue volume can be quantified by integrating the voxel values within a volume of interest. The aim of the automatic quantification is therefore restricted to locating the volume of interest, e.g. the visceral compartment, rather than separating adipose from non-adipose tissue, which would be necessary if using non-quantitative images (Figure 1).

The anatomical compartments for adipose tissue were labelled: Total Adipose Tissue (TAT), Abdominal Subcutaneous Adipose Tissue (ASAT: fat superficial to the deep fascia of the muscles of the trunk), and Visceral Adipose Tissue (VAT: fat confined to the abdominal cavity deep to the muscles of the trunk) (Figure 2). The semi-automated segmentation of the ASAT and VAT compartments was performed using multi-atlas non-rigid registration. The atlas database consisted of \( N = 11 \) (4 females and 7 males) manually segmented whole body datasets with a BMI range of 19.7 to 32.3 kg/m\(^2\), a VAT range of 1.05 to 7.43 L and a mean VAT volume of 3.60±1.93 L, where the VAT and ASAT compartments were pre-defined.

The water image of each atlas was non-rigidly co-registered to the subject’s water image. The resulting displacement field was then applied to the visceral compartment mask and water images of the atlas. This was repeated for all atlases \( N \) in the database, resulting in \( N \) segmentations of VAT and ASAT compartments. The four atlases with the highest normalized cross correlation between the subject’s and the atlases’ water images within the image planes covering the abdomen were then selected. Finally, the VAT and ASAT masks were defined as the regions were at least two of the four selected atlases agreed on the classification as VAT or ASAT.
After the automatic segmentation, a manual quality assurance inspection by a trained image analyst employed by AMRA was made on the resulting VAT and ASAT masks. The operator could then locally adjust the resulting segmentations in case of atypical morphologies. The operators had two weeks training and at least 6 months experience of using this tool. Finally, the amounts of VAT and ASAT were automatically quantified by integration of the calibrated fat image volume within the VAT and ASAT masks respectively.

The TAT was completely automatically obtained by integrating the fat image volume inside a body mask. The body mask was obtained by thresholding the sum of the water and fat images at 0.5 (50 % of pure tissue signal) and then closing all cavities.

**Manual segmentation technique**

Manual segmentation of visceral fat was performed independently, using SliceOmatic™ 5.0 (Tomovision Inc., Quebec, Canada) by two radiology trainees each with 4 years experience (DN,CKM) under the direct supervision of a senior radiologist (18-years’ experience (AT)). The visceral compartment was defined on both semi-automated and manual segmentation to include the entire intraperitoneal fat volume including the pelvis.

**Statistics**

Statistical analyses were performed using R statistical software (Version 3.0.1 including psych package) (30). All but one of the datasets disproved the Shapiro-Wilks test for normality and therefore the descriptive statistics are presented as non-parametric data. Coefficients of variation (CV) were calculated as the square root of the mean within-group one-way variance (ANOVA) divided by the grand mean and expressed as a percentage. The absolute two-way intraclass correlation coefficient (ICC) using the same raters for all subjects was calculated for both semi-automated segmentation and manual segmentation. Test-retest reliability and inter-method comparison was performed using Bland–Altman plots to calculate the 95% limits of agreement and the Coefficient of Repeatability (CR), also known as the Smallest Real Difference (SRD). The CR or SRD is calculated as 1.96 x SD of the differences between paired observations (31, 32). A pairwise Wilcoxon signed Rank test with Holm correction for multiplicity was used to compare semi-automated with manual segmentation.

**RESULTS**

Thirty participants were recruited, 15 male and 15 female. The mean age was 39.8 years (range 26–56).
The median TAT volumes were 30.88 L (Interquartile range (IQR) 23.7–40.6 L) for the first MR acquisition and 30.65 L (IQR 23.9–40.4 L) for the second. The ICC for first and second measures was perfect with ICC = 1.0. The mean difference between the measurements was -0.04L with 95% limits of agreement of -0.67 L and 0.6 L, a CV of 0.79% and a CR of 0.49 L.

The median ASAT volumes were 7.61 L (IQR 4.98–10.10 L) for the first MR acquisition and 7.60 L (IQR 4.97–9.97 L) for the second with a perfect ICC = 1.0. The mean difference in measurements was -0.05 with 95% limits of agreement of -0.55 L and 0.64 L, a CV of 2.98% and a CR of 0.46 L.

The median semi-automated VAT volumes were 2.02L (IQR1.20–3.18) for the first MR acquisition and 2.03 L (IQR 1.21–3.18 L) for the second, again with perfect ICC = 1.0. The mean difference in measurements was -0.04 with 95% limits of agreement of -0.12 L, 0.13 L, a CV of 1.80% and a CR of 0.09 L.

The automated intensity inhomogeneity correction and calibration of the fat image volume took approximately 10 min and the atlas-based segmentation took approximately 7 min per atlas on a standard PC. The time spent on this manual inspection and interaction was typically less than 3 minutes per subject and compartment.

The median manually segmented volumes of VAT were 2.55 L (IQR 1.78–3.65 L) for observer 1 and 2.39 L (IQR 1.76–3.57 L) for observer 2. Intraclass correlation was nearly perfect with ICC = 0.99 (95% confidence intervals 0.97, 0.99). The mean difference between observers was 0.13 L with 95% level of agreement -0.51 L, 0.77 L, a CV of 6.30% and a CR of 0.39 L. (Tables 2 & 3, Figure 3). The time taken for each observer to manually segment visceral fat was approximately 40 minutes.

A four-way comparison of the two semi-automated and two manual VAT datasets demonstrated differences in the mean of between 0.49 L and 0.63 L, and the median of 0.37 L and 0.53 L.There was no significant difference between the semi-automated and manually segmented volumes (P=0.73). There was “almost perfect” intraclass correlation of ICC = 0.97 (95% CI: 0.85–0.99). The CV varied from 7.0–10.5% and the CR from 0.49–0.59 L (Table 4, Figure 4).
DISCUSSION

The results of this study suggest that quantification of whole body fat, abdominal subcutaneous fat, and visceral fat can be reliably performed across a range of body masses, from normal to obese, on a wide bore 3T MR system. There are challenges in wide bore MR imaging, body imaging at 3T and in MR imaging of large patients. Despite this the coefficients of variation (CV) and reliability (CR) between first and second acquisitions suggest that semi-automated whole body, abdominal subcutaneous and visceral fat volume measurements may be suited to longitudinal studies.

The intraclass correlation coefficients (ICC) demonstrated “perfect” agreement for all the semi-automated test-retest measures and “almost perfect” agreement for the manually segmented VAT (33). While ICC is commonly used to measure the level of agreement in test-retest reliability studies these results highlight the limitations of the test. ICC is a measure of linear correlation between the two datasets that accounts for consistency within subjects as well as changes in mean performance and therefore is better suited to correlation of repeated measurements than conventional tests of correlation such as Pearson’s Product Moment Correlation coefficient (34). However ICC is also affected by the characteristics of the data distribution and in particular large ranges can produce high correlations and this may account for the very high ICC in this study (35). The perfect ICC of 1.0 for all semi-automated measurements does not reflect the true test-retest variability.

The test-retest reliability is better reflected by the CV and CR. The CV provides a standardized deviation for the spread of the differences between first and second datasets related to the mean of the measurements. This allows comparison between the test-retest reliabilities of the different compartments. The smallest CV for semi-automated measures is for TAT at less than 1% and the largest is for ASAT at just under 3%. These results are similar to previously published data of 2.08% for total trunk adipose and 2.11% for ASAT, at 3.0T on a flared-bore system with a maximal 50cm field of view (21), and 1.01% for TAT (36)[Kullberg], and 2.25%–3% for ASAT at at 1.5T on conventional bores (22, 23). The CV for manual segmentation was 3.5 times larger than that for the semi-automated segmentation.

Whereas the CV provides a relative measure of variance the CR provides an absolute measure of the spread of differences between first and second acquisitions and for a number of authors this is the optimal measure of test-retest reliability (25,(37). For TAT the CR is 0.49, which means that for a population with a mean TAT volume of approximately 30 litres the Smallest Real Difference (SRD) is about 500mL. In other words 95% of all differences between first and second measures will be less than 500mL—approximately 1.6% of the mean. The CR of 0.46 for ASAT constitutes a little less than 6% of the mean, and the CR of 0.09 for VAT constitutes about 3.5% of the mean. In comparison the CR for the
Test-retest reliability of manually segmented VAT was 0.39, which was 12.5% of the mean. Therefore despite uniformly high ICC, semi-automated quantification on a 3.0T wide bore MR machine appears to be quicker and more reliable than the manual technique.

The primary aim of the study was to measure test-retest reliability and not validate atlas based segmentation, which has been reported elsewhere (38)(39). Therefore we chose to manually segment only the VAT because we predicted that the proportionate measurement errors were going to be greatest in this compartment. This is because the compartment is relatively small compared to TAT and ASAT, the shape of the compartment is more complex and because there is more potential for the shape to change from first to second acquisition as the bowel moves. The argument was that this would give us the greatest discrepancy in an inter-method comparison, which was the secondary aim of the study, and that this is most likely to be a focus of future longitudinal studies.

The results presented here are in line with previously published validation studies of the calibration of fat images to produce quantitative signal (40)34) and validation of the method for whole body fat quantification against air displacement plethysmography (42).

Test-retest reliability data for automated quantification of whole body fat has previously been reported by Wald et al. with coefficients of variation of between 3% and 6% for repeated measures of VAT and subcutaneous adipose respectively (43). There are a number of important differences between Wald’s study and ours. The major technical difference is that Wald’s technique relies on spatial segmentation of fat voxels whereas ours uses quantitative data irrespective of anatomical location. This means that the technique described in this paper should be less susceptible to partial volume effects, particularly in complex areas such as the abdomen, although this has not been assessed in this study. Wald and colleagues also removed the data from the arms, because they were restricted by the field of view and the bore of the magnet. Our study was performed on a wide bore MR unit with a horizontal diameter of 70cm so that the data includes both arms and therefore a more complete measure of TAT. However in some patients there was signal loss at the periphery of the field of view that meant that data was lost from the lateral surface of the arms. Therefore while we would argue that this brings us closer to a true TAT it is not quite complete. We do not know how much of a difference including the arms makes to the final fat volume as we have not performed quantification without including the arms and our cohort is not comparable to that of Wald et al.

In addition to the larger bore diameter we also used a higher magnetic field strength. It is more challenging to maintain field homogeneity in a wide bore machine and to control dielectric effects at 3T and therefore these were both potential problems with using our system. However, our results suggest that our methodology may be more reproducible than previously published techniques. Using a comparable measure our coefficient of variation for TAT was 0.79% and 1.8% for VAT. This is arguably better than previous VAT reliability
measures despite the challenges of the 3T field and the wide bore and similar to previous measure of reliability for TAT.

The final difference in these two studies is the size of the subjects. The mean TAT for our study was 31.74L (SD 13.6L) compared to a mean of 16.4L (SD 3.3L) in Wald’s study. We deliberately recruited 10 volunteers with a BMI over more than 30 kg/m^2 to test whether the reliability of the technique could be maintained in larger subjects. It would be reasonable to suspect that the greater the distortion to the B0 magnetic field the greater the test-retest variability might be. However the Bland-Altman plot for TAT (Figure 6) demonstrates no funnelling of the data points indicating that the reliability is maintained in obese patients. There was a single subject with a BMI of over 40 kg/m^2 that placed outside the 95% limits of agreement, which suggests that reliability may deteriorate above a BMI of 35 kg/m^2.

There was no significant difference between the manual and automatic visceral fat quantification. However there is a difference in the variability around these median measurements with larger CV and CR for the manual compared to the automated data.

The manual segmentation technique used in our study (SliceOmatic™) has limitations that arise from the method of grouping pixels according to signal intensity thresholds. This can result in inaccurate inclusion of intramuscular and subcutaneous fat and fat within intra-abdominal organs. The semi-automated thresholding can include these additional areas of fat, which then require the operator to manually adjust the semi-automated segmentation; this is likely to account for most of the inter-observer variability. The CV for manual test-retest reliability of VAT was 6.3%, which compares favourably with previously published CVs of VAT of between 5–13% for a range of manual segmentation software packages (including SliceOmatic™) in a side-by-side comparison (44). This suggests that the variability measured in manual segmentation in our study derives from the technique and is not specific to our operators.

A number of study design limitations should be considered. The selection of subjects was limited to three well-defined categories: normal, overweight and obese. No underweight subjects were included in the study and even the increased diameter of the widebore magnet will provide a limit to the girth of larger patients who could be included. This means that the results are not necessarily applicable to all populations. There could also be differences in the reliability of fat quantification between obese men and women because the characteristic fat distribution—peripheral versus central—may have a differential influence on field inhomogeneities. The number of male and female subjects in this study did not allow us to perform this comparison. The test-retest component of the study was limited to two examinations on the same day relatively close together and therefore probably represents the best available reliability. Reliability may decrease as the time
between examinations increases and if different operators are scheduled for each acquisition.

In conclusion completely semi-automated whole body fat volume quantification and semi-automated compartmental fat volume quantification can be achieved in less than 15 minutes on a 3T wide bore MR system with small CV and acceptable CR in normal volunteers and subjects with a BMI of more than 30 kg/m². Comparison with manual segmentation of VAT suggests that this semi-automated technique is quicker, more reliable and produces equivalent median measurements.
REFERENCES

1. WHO | Obesity and overweight [http://www.who.int/mediacentre/factsheets/fs311/en/]


### Table 1

<table>
<thead>
<tr>
<th>BMI category (kg/m²)</th>
<th>Number of volunteers</th>
<th>Height (cm)</th>
<th>Weight (kg)</th>
<th>BMI (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–25</td>
<td>10</td>
<td>172 (18)</td>
<td>65.6 (15.2)</td>
<td>22.1 (2.7)</td>
</tr>
<tr>
<td>25–30</td>
<td>10</td>
<td>170 (16)</td>
<td>77.2 (18.2)</td>
<td>26.6 (2.3)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>10</td>
<td>170 (16)</td>
<td>97.6 (30)</td>
<td>33.8 (10)</td>
</tr>
<tr>
<td>All male</td>
<td>15</td>
<td>176 (7)</td>
<td>85 (27.3)</td>
<td>27.5 (10)</td>
</tr>
<tr>
<td>All female</td>
<td>15</td>
<td>165 (10)</td>
<td>75.3 (38.6)</td>
<td>27.5 (13)</td>
</tr>
<tr>
<td>All volunteers</td>
<td>30</td>
<td>171 (17)</td>
<td>80 (34)</td>
<td>27.5 (11.4)</td>
</tr>
</tbody>
</table>

Table summarizing the demographics of the study population (mean ±2 standard deviations).
Table 2

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Episode</th>
<th>Median (L)</th>
<th>Interquartile range (L)</th>
<th>Mean difference (L)</th>
<th>95% level of agreement (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total adipose tissue (TAT)*</td>
<td>1</td>
<td>30.88</td>
<td>(23.7–40.6)</td>
<td>-0.04</td>
<td>-0.67, 0.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>30.65</td>
<td>(23.9–40.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal subcutaneous adipose tissue (ASAT)*</td>
<td>1</td>
<td>7.61</td>
<td>(4.98–10.10)</td>
<td>0.05</td>
<td>-0.55, 0.64</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7.60</td>
<td>(4.97–9.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visceral adipose tissue (VAT)*</td>
<td>1</td>
<td>2.02</td>
<td>(1.20–3.18)</td>
<td>-0.04</td>
<td>-0.12, 0.13</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.03</td>
<td>(1.21–3.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visceral adipose tissue (VAT)</td>
<td>1</td>
<td>2.55</td>
<td>(1.78–3.65)</td>
<td>0.13</td>
<td>-0.51, 0.77</td>
</tr>
<tr>
<td>Manual</td>
<td>2</td>
<td>2.39</td>
<td>(1.76–3.57)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Semi-automated  
† Derived from Bland-Altman plots

Table summarizing the volumes of segmented adipose tissue with reliability and method comparison statistics.
Table 3

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Coefficient of variation (%)</th>
<th>ICC (95% CI)</th>
<th>Mean Volume (L)</th>
<th>Coefficient of Repeatability (CoR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total adipose tissue (TAT)</td>
<td>0.79</td>
<td>1.0</td>
<td>31.74</td>
<td>0.49</td>
</tr>
<tr>
<td>Abdominal subcutaneous adipose tissue (ASAT)</td>
<td>2.98</td>
<td>1.0</td>
<td>7.92</td>
<td>0.46</td>
</tr>
<tr>
<td>Visceral adipose tissue (VAT) Semi-automated</td>
<td>1.80</td>
<td>1.0</td>
<td>2.56</td>
<td>0.09</td>
</tr>
<tr>
<td>Visceral adipose tissue (VAT) Manual</td>
<td>6.33</td>
<td>0.99</td>
<td>3.12</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Table summarizing measures of variation and repeatability between first and second measures (test-retest). The ICCs without 95% confidence intervals presented demonstrated perfect intraclass correlation.
Table 4

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Mean of the differences(^*) (L)</th>
<th>Difference in medians (L)</th>
<th>Significance (P)(^*)</th>
<th>CV (%)</th>
<th>CR (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated 1</td>
<td>-0.62 (-0.73, -0.52)</td>
<td>0.53</td>
<td>0.73</td>
<td>10.3</td>
<td>0.58</td>
</tr>
<tr>
<td>Observer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated 1</td>
<td>-0.49 (-0.58,-0.4)</td>
<td>0.37</td>
<td>0.73</td>
<td>7.4</td>
<td>0.49</td>
</tr>
<tr>
<td>Observer 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated 2</td>
<td>-0.63 (-0.74, -0.52)</td>
<td>0.52</td>
<td>0.73</td>
<td>10.5</td>
<td>0.59</td>
</tr>
<tr>
<td>Observer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated 2</td>
<td>-0.49 (-0.59, -0.41)</td>
<td>0.37</td>
<td>0.73</td>
<td>7.0</td>
<td>0.49</td>
</tr>
<tr>
<td>Observer 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^*\) Mean of differences with 95\% confidence intervals for the mean

\(^*\) Pairwise Wilcoxon Signed Rank with Holm adjustment for multiplicity

Table summarizing comparison of manual and semi-automated visceral fat segmentation.
Diagram representing segmentation post-processing. The water images of the Atlas Database (upper left) are non-rigidly registered to the Target water image (bottom left) producing a Displacement field. This is then applied on the Labels in the Atlas Database to produce a Deformed label field (centre). The deformed labels are combined into a Probability map which is thresholded to produce labels for the targets (right). Finally, the labels are applied onto the target fat image, producing a segmented and labelled target image (bottom right).
Figure 2

Color illustration of fat quantification. From left to right: (A) Raw fat image; (B) Quantitative fat image after calibration (TAT); (C) Semi-automated classification of visceral fat (green); (D) Final result after manual interaction (VAT). The differences between C and D are very small illustrating that the manual interaction comprises only minimal changes in quantification.
Bland-Altman plots demonstrating the mean versus difference in observations and 95% limits of agreement for test-retest reliability of semi-automated total adipose tissue (TAT) and abdominal subcutaneous adipose tissue (ASAT) volume measurements.
Figure 4

Bland-Altman plots demonstrating the mean-difference plots of observations, and 95% limits of agreement, for method comparison of semi-automated versus manually quantified visceral fat volumes.