MRI adipose tissue and muscle composition analysis: a review of automation techniques

Magnus Borga

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-149809

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

Original publication available at:
https://doi.org/10.1259/bjr.20180252

Copyright: British Institute of Radiology
http://www.bir.org.uk/
MRI adipose tissue and muscle composition analysis – a review of automation techniques

Magnus Borga, PhD

Department of Biomedical Engineering, and
Centre for Medical Image Science and Visualization (CMIV)
Linköping University
Sweden

Abstract

Magnetic resonance imaging (MRI) is becoming more frequently used in studies involving measurements of adipose tissue (AT) and volume and composition of skeletal muscles. The large amount of data generated by MRI calls for automated analysis methods. This review article presents a summary of automated and semi-automated techniques published between 2013 and 2017. Technical aspects and clinical applications for MRI-based AT and muscle composition analysis are discussed based on recently published studies. The conclusion is that very few clinical studies have used highly automated analysis methods, despite the rapidly increasing use of MRI for body composition analysis. Possible reasons for this are that the availability of highly automated methods has been limited for non-imaging experts, and also that there is a limited number of studies investigating the reproducibility of automated methods for MRI-based body composition analysis.
INTRODUCTION

Tomographic techniques, i.e. computed tomography (CT) and magnetic resonance imaging (MRI), are recognized as state-of-the-art methods for body composition analysis, in particular for quantification of compartmental volumes of adipose tissue (AT)\(^1\) and muscles\(^2\). High soft-tissue contrast, in combination with increasing availability and absence of ionizing radiation, makes MRI more frequently the preferred method of choice. In several large population studies, such as the UK Biobank\(^3\), the German National Cohort\(^4\), the KORA-MRI study\(^5\), the Netherlands Epidemiology of Obesity Study\(^6\) and the Dallas Heart Study\(^7\), MR images are acquired to enable advanced body composition analysis. For example, in the UK Biobank Imaging Study, MR image volumes from 100,000 subjects are being collected, each volume containing 332 axial slices covering 1.1 m of the abdomen and upper legs. Although the overall aim of these studies is to collect data for future research, and the analysis methods have not been prescribed in the study design, it is quite obvious that automated methods are required in order to analyse such large amounts of imaging data.

The direct volumetric measurements of single muscles or fat compartments obtained by MRI enable much higher accuracy compared to indirect measurements such as anthropometric measures. For example, visceral AT (VAT) can vary significantly between people with identical body mass index (BMI) or waist circumference\(^8, 9\). It is also well-known that the metabolic risk related to fat accumulation is strongly dependent on its distribution\(^10, 11\). In particular, large amounts of VAT are related to increased risk for cardiac disease\(^10, 12, 13\), type-2 diabetes (T2D)\(^14, 15\), liver disease\(^16\) and cancer\(^17, 18\). Furthermore, increased muscle fat infiltration has been associated with reduced mobility\(^19\), increased risk for T2D\(^20\) and higher mortality in patients with liver cirrhosis\(^21\).

While two-dimensional projections of the body using dual-energy x-ray absorptiometry (DXA) have high agreement with volumetric measurements using MRI for whole-body fat and lean tissue quantification, compartmental measurements such as VAT have a relatively low agreement\(^22, 23\). Also, while CT technically has the same capability as MRI to acquire complete 3-dimensional image volumes, body composition analysis with CT is commonly restricted to one or a limited number of slices in order to reduce the radiation exposure. But the use of 2-dimensional area measures from one or a limited set of slices as a proxy for volume reduces the precision compared to volumetric measurements when measuring abdominal AT distribution\(^24, 25\). The challenge, of course, with 3-dimensional volumetric body composition analysis is the huge amount of data needed to be analysed. Completely manual analysis of full tomographic image volumes covering a large part of the body is extremely time consuming and hence, generally not feasible except in very small studies. This has generated a demand for automated methods for quantifying AT and muscle composition in 3-dimensional MRI images. The feasibility of using a highly automated analysis method in a large population study was, for example, demonstrated by West et al.\(^26\), quantifying VAT, abdominal subcutaneous adipose tissue (ASAT) and thigh muscles on 3,000 subjects from the UK Biobank Imaging Study.

Another advantage with automated analysis methods is their potentially higher precision because of reduced or eliminated dependency on human operator variability. For example, in a study by Newman et al.\(^27\), significantly higher test-retest repeatability of VAT quantification was achieved with a highly automated method compared to using manual analysis (1.8 % vs. 6.3 % coefficient of variation).

The aim of this review was to give an overview of automated and semi-automated methods for MRI-based body composition analysis by presenting a summary and discussion of methods published between 2013 and 2017. Recent clinical applications of automated
METHODS FOR MRI-BASED BODY COMPOSITION ANALYSIS

Any method for body composition analysis using MRI can be divided into three main steps: image acquisition, image segmentation, and tissue quantification. All three steps affect the results of a body composition analysis, and although this review focuses on automation of image post processing, and not image acquisition, relevant image acquisition methods are also discussed here since the choice of acquisition method has an impact on the subsequent analysis.

Image Acquisition

MRI is often not used in a quantitative way, i.e. the image intensity values do not directly reflect a physical property of the imaged object (as opposed to CT, which is calibrated against the Hounsfield scale). Imperfections in the MRI system, as well as interactions between the imaged object and the electro-magnetic field, cause the sensitivity and hence the image intensity scale to vary over the image. In such cases, the interpretation of the images has to rely solely on the visual appearance of the contrast between tissues in the image. This limits the possibilities of accurate quantification of the image data. Still, both T1- and T2-weighted images have good contrast between water and fat. This means that lean tissue (LT) and AT can be segmented from each other (as long as the image resolution is high enough), enabling quantification of geometric properties such as area or volume of different tissue compartments using many standard T1- or T2-weighted imaging protocols.

In the context of AT and muscle analysis, the fat-water (FW) separated imaging (a.k.a. "Dixon imaging") techniques are particularly useful. FW-separated imaging is based on gradient recalled echo (GRE) imaging, which uses the chemical shift between the resonance frequencies of protons bound in water and in fat. This is the same effect which is used in magnetic resonance spectroscopy (MRS). Fat-water separated imaging can be seen as a special case of MRS, where the spectral resolution has been sacrificed in favour for spatial resolution. In FW-MRI, two or more echoes are acquired after each excitation pulse, where the fat and water signals' relative phase differ due to their chemical shifts. Besides the separation of the fat and water components of the MR signal, measurements of multiple echoes enable estimation of several unknown confounding factors in the signal equations, such as T1 and T2*, which can then be corrected for. A well-known example of FW-MRI is the IDEAL reconstruction method.

Besides the excellent fat-water contrast enabled by FW-MRI, it also enables quantitative fat imaging. One common method of achieving quantitative fat images is to compute the fat fraction (FF), which is the fat signal divided by the sum of the fat and water signals. A well-known version of FF is proton density fat fraction (PDFF), which can be obtained by using proton-density weighted FW separated imaging. Another example is fat-referenced MRI, which measures the fat signal in a given voxel in relation to the fat signal in pure adipose tissue. An example of a whole-body fat image that has been calibrated using fat-referenced MRI is shown in Figure 1.
Also, brown adipose tissue (BAT) imaging using MRI requires quantitative imaging. BAT has most commonly been imaged using positron emission tomography (PET) but, more recently, dual energy computed tomography (DECT) and MRI have also been used to detect and quantify BAT. Quantitative fat-water separated MRI can be used for identification and characterization of BAT by its lower fat content and higher water content compared to white AT. Quantitative fat-water separated MRI can be used for identification and characterization of BAT by its lower fat content and higher water content compared to white AT. A challenge when using FW-MRI to detect BAT is that it requires a high resolution in order to separate the BAT from partial volume effects in the interfaces between white AT and LT, while maintaining sufficiently high signal-to-noise ratio. Another contrast mechanism which can be used to characterize BAT is T2* relaxation. T2* mapping can be combined with FW separation using multi-echo chemical-shift imaging. In addition to detecting the presence of BAT, the activation of BAT can be detected by changes in T2* due to an increase in blood deoxyhemoglobin levels, which is caused by increased oxygen consumption in active BAT.

Image Segmentation

There is often not a clear distinction made between tissue classification and compartmental segmentation, but usually the tissue classification is done before the segmentation into compartments. Classification, in this context, refers to the labelling of each voxel into a tissue class, e.g. LT, AT or background. Segmentation, on the other hand, refers to distinguishing between different compartments containing the same type of tissue (e.g. VAT from abdominal subcutaneous AT (ASAT) or one muscle from another) by labelling or delineating relevant anatomical regions. If the image is quantitative, i.e. calibrated in relation to a defined grey scale, tissue classification is rather straightforward as it can be based on thresholding the image intensity; in particular in fat-water separated images. In non-quantitative imaging, the tissue classification can be facilitated by reducing the image inhomogeneities present in non-quantitative MR images. Such intensity inhomogeneities are caused by imperfections in the image acquisition system, such as inhomogeneities in the magnetic field and variations in the coil sensitivities. This causes variations in the intensity scale over the image. A number of methods have been proposed for correcting MR images for such inhomogeneities, and a comprehensive review and classification of such methods is given by Vovk et al. They divided the methods generally into prospective and retrospective methods, where the prospective methods try to solve the problem during the image acquisition while retrospective methods use image analysis postprocessing to estimate and correct for the inhomogeneities. An important advantage with retrospective methods is that they can also correct for inhomogeneities induced by the body in the scanner, since they make very few assumptions about the physical sources of the inhomogeneities. The most common assumption being made is that the inhomogeneities can be modelled as a spatially slowly varying bias field and the different methods use different image processing methods to estimate and remove this bias field. There have also been other methods published after the review by Vovk et al., e.g. Consistent Intensity Inhomogeneity Correction (CIIC) which is based on multi-scale normalized averaging.

A common way to classify voxels into different tissues when the image is non-quantitative and the grey scale is undefined, is k-means clustering or fuzzy c-means clustering. In both methods, a predefined number \( k \) (or \( c \)) of clusters is used to categorize the data. In the case of body composition analysis, \( k \) or \( c \) are usually set to 3, indicating the classes AT, LT and background. The difference between the two methods is that k-means uses crisp (mutually exclusive) class memberships, while fuzzy c-means uses continuous ("fuzzy") class memberships.
membership values. Note that classification is actually not required when quantitative images are used, since each pixel or voxel then implicitly contains a class-membership value.

Proposed segmentation methods have been more or less automated, ranging from computer-aided manual drawing tools to fully automated segmentation algorithms. Less automated methods are often more generally applicable to different segmentation tasks, but at the cost of more time-demanding (and therefore more expensive) manual work. In 2-dimensional images, a common segmentation method is active contours (a.k.a. "snakes") where a deformable contour is fitted to a structure in the image. This class of methods works by minimizing an energy function with one term depending on the image (e.g. the image gradient) and a regularization term depending on the curvature of the contour. While there are 3-dimensional extensions of active contours ("balloons") , a more common class of 3-dimensional segmentation methods in this context is multi-atlas segmentation, where several anatomical atlases with pre-defined compartments are fitted to the image and whose anatomical definitions are combined using a voting scheme or expectation maximization. While atlas-based segmentation has mainly been used for brain image segmentation, they have recently also been applied to abdominal and whole-body image analysis. Beside active contours and multi-atlas segmentation, there are a multitude of different segmentation methods in the literature, and even more ways of combining such methods. These two segmentation methods were, however, found to be the most commonly used in the studies covered in this review. Atlas-based methods are, in principle, easier to generalize to other compartments compared to methods based on rule-based morphological operations that rely on specific anatomical assumptions.

Few fully automated segmentation methods have been extensively evaluated on different datasets with different properties (e.g. from different scanner models and different field strengths). Notable exceptions are the work by Addeman et al., where a fully automated algorithm for segmentation of abdominal AT into VAT and ASAT was evaluated on data from scanners of three different brands and two different field strengths, and the work by Karlsson et al., where fully automated segmentation of muscle groups from whole-body MRI was evaluated on data from two scanners with different field strengths.

A fully automated segmentation tool does not necessarily allow the user to modify the result, which obviously limits the range of images that can successfully be segmented. Such methods may also face a challenge from a regulatory perspective different to methods where the result is controlled by a human operator. One class of methods, which can be referred to as supervised segmentation (following the terminology by Hu et al.), combines the efficiency of fully automated methods with the flexibility and control enabled by manual interaction.

The difference between a supervised automated tool and a semi-automated tool is that the latter can never do the complete segmentation without human interaction, while the supervised tool, at least in principle, can always autonomously perform a segmentation, albeit one that may need to be manually adjusted to give the desired result.

Segmentation methods can also be divided into two-dimensional (2D) and three-dimensional (3D) methods. 2D methods operate on one image slice at a time and are often used to segment one or a limited number of slices. Analysing only one or a few image slices with a manual or semi-automated tool is, of course, much less time consuming than analysing all slices in a complete volume, and many studies using MRI for body composition analysis have used one or a limited set of two-dimensional slices, mostly due to the lack of efficient image analysis tools for handling three-dimensional image segmentation. However, software tools for slice-wise semi-automated segmentation took on average more than 10 minutes per slice for a trained expert, which limits their use to small studies and to measurements of a limited set of slices. The rationale of using area as a proxy for volume is based on modelling the body as
a cylinder. While such a model might be reasonable for the extremities, it is not well suited for the abdominal compartment, where single-slice analysis cannot accurately measure intra-abdominal AT\textsuperscript{24}. When the tissue area in one or a few slices is used as proxy for VAT and ASAT volume, the location of the slices is critical. While such area measurements can have a good correlation with the absolute volume, single-slice imaging does not have the accuracy required to measure VAT and ASAT changes in an intervention study\textsuperscript{25}. For automated segmentation, 3D methods should (at least theoretically) be more powerful than 2D methods, since the connectivity of a non-convex 3D object may be lost when viewed as a 2D slice. Also, information from neighbouring slices (which is inherent in 3D methods) increases the redundancy and, hence can alleviate detection of significant structures and thereby support the segmentation in noisy images.

**Tissue Quantification**

In non-quantitative imaging, the actual tissue quantification is usually performed by simply counting the number of voxels of each tissue class within each compartment and multiplying with the volume of a voxel. Such methods will be referred to as *discrete* methods. When using quantitative images, on the other hand, the amount of fat in a segmented compartment can be quantified by integrating the fat signal within that compartment, multiplied by the voxel volume. Such methods will be referred to as *continuous* methods. While continuous methods do not require a classification between AT and LT, they usually include a classification of soft tissue vs. background (i.e. air and MR-invisible tissue) in order to remove noise contributions from the background.

The disadvantage with discrete methods is that partial volume effects, i.e. voxels with mixed content, will lead to a bias in the volume estimates\textsuperscript{71, 72}; a bias that will change with the resolution, hence limiting the reproducibility across different scanning protocols. This effect is illustrated in Figure 2 where an original quantitative fat image (top left) has been sub-sampled a number of steps (left column). A discretisation of the images on each resolution has been made using a threshold of 0.5 (second column). The diagram to the right shows, for different resolutions, the estimated AT area using a continuous method (solid line) and a discrete method with different thresholds (dashed lines). In the continuous method, all pixel values in the quantitative fat image are summed and then multiplied with the pixel area. The discrete method applies a threshold on the fat image and then multiplies the number of pixels above the threshold by the pixel area. A brief explanation of this resolution-dependent bias is as follows: For large objects, partial volume effects will only occur at the tissue-background interface. Peripheral voxels covering the object to a fraction higher than the threshold will be classified as object and, hence, over-estimate the volume, while voxels covering the object less than the threshold will be considered as background, hence under-estimating the volume. For a threshold of 0.5, these errors will, on average, cancel each other out, hence giving an un-biased estimate\textsuperscript{71} (assuming a symmetric noise distribution with zero mean). But objects (or parts of objects) that are smaller than half the voxel size in any dimension (e.g. a thin sheet of tissue) will all be classified as background and lost. Hence, a discrete classification of voxels will under-estimate the object volume, and this bias will increase with decreasing image resolution. The more irregularly shaped and thin the tissues are, the more prominent this effect will be, and in the case of muscle fat infiltration, the effect can be quite significant\textsuperscript{72}. This will make discrete methods difficult to use in multi-centre studies, where the scanner model and scanning protocol may differ.

Continuous methods are less sensitive to partial volume effects, which makes them less prone to the resolution-dependant bias discussed above. But, perhaps more importantly, in contrast
to discrete methods, continuous methods enable measurements of diffuse infiltration of thin AT structures and ectopic fat in non-adipose tissue, such as liver and muscles, since these methods can quantify weak fat signals that would not reach any reasonable threshold for classifying a voxel as AT.

**Accuracy and Precision**

Few studies have investigated repeatability and reproducibility. Reproducibility measures how well the measurement can be reproduced with different scanners while repeatability measures how well the analysis of a subject scanned in the same scanner agrees with the analysis of a second scan, or between two or more analyses made by the same or different operators. Most studies have only evaluated accuracy of the segmentation against manual segmentation. But from a clinical perspective, repeatability and reproducibility are at least as important as accuracy. Evaluation of the segmentation alone does not address the quantitative properties of the complete imaging chain. For example, sensitivity to differences in scanning parameters such as image resolution is not reflected by comparison to manual segmentation, and different methods may be more or less sensitive to partial volume effects, which are directly related to image resolution.71,72

**REVIEW OF RECENT LITERATURE**

Starting in the early 1990s, a wide range of papers on AT quantification using MRI have been published. An excellent review of methods for segmentation of AT was recently presented by Hu et al. The present review is constrained to the last five years (2013 - 2017), during which a range of papers on more or less automated methods for quantification of abdominal AT, BAT and muscles have been published. A search was made on PubMed for studies using MRI with automated or semi-automated methods for quantifying fat or muscles. 34 publications are summarized in Table S1 (supplementary material) with respect to measurements; image pre-processing (e.g. inhomogeneity correction); segmentation method; if the image analysis was performed slice-wise in 2D or if a full 3D processing method was used; if a complete volume measure was obtained or if an area-measure was used as a proxy; the level of automation; the type of validation reported, and the repeatability if reported.

**Quantitative Imaging and Inhomogeneity Correction**

More than half of the 34 papers used either continuous methods or applied some kind of inhomogeneity correction to aid the segmentation. Seven papers used continuous quantification using based on quantitative MRI, either PDFF or fat-referenced MRI. Eleven of the papers that used discrete quantification methods used inhomogeneity correction before AT quantification or muscle analysis. The remaining studies used 2-dimensional area measures as proxy for the volume. Active contours were used for segmentation in most of the 2-D methods, while in the studies...
using 3-dimensional segmentation, multi atlas-based segmentation was the most common method, used in eight of the studies.

Level of Automation

17 studies used fully automated methods for AT quantification and muscle measurements. One study used completely manual segmentation of baseline images and automated non-rigid registration for transferring the baseline segmentations to follow-up images. The remaining studies used semi-automated or supervised segmentation methods.

Validation

Most of the studies reported accuracy in terms of agreement with manual segmentation, but only 14 of them reported precision in terms of repeatability. All of the studies that reported precision showed very good repeatability, with coefficients of variation of up to a few percent and the intra-class correlation (ICC) was very close to one in most of the studies.

The only study that investigated the between-scanner reproducibility was presented in the paper by Karlsson et al., who compared fully automated quantification of 10 different muscle groups in data acquired on a 1.5 T and a 3 T scanner with excellent agreement. Also, Addeman et al. addressed the issue of multi-centre studies, showing their method’s ability to analyse data from three different cohorts using different scanners and field strengths. They did, however, not evaluate between-scanner reproducibility since the different scanners were used on different cohorts.

Availability

Most of the methods used in these studies are not readily accessible for other researchers and non-image analysis experts. Three exceptions are FATCALC; AMRA Researcher, which was used in six of the studies listed above, and SliceOmatic used in one of the studies.

FATCALC is an add-on to ImageJ, which is a free image processing software originally developed by NIH for analysis of microscopy images. FATCALC is a discrete 2-dimensional quantification method for VAT and ASAT. It works on images acquired using FW-MRI, using 2D analysis of individual slices and fuzzy c-means clustering to classify AT pixels in the fat image. This is followed by morphological operations (erosions and dilations) of the binary classification result in order to separate VAT from ASAT. Finally, fat pixels belonging to the arms, abdominal muscles and paravertebral fat are removed. The complete process is fully automated.

AMRA Researcher (AMRA Medical AB, Linköping, Sweden) is a cloud-based analysis service, in which the volumes of VAT, ASAT and different muscle groups, as well as muscle fat infiltration and liver fat are quantified. It uses a continuous quantification method based on quantitative fat-referenced MRI and 3D multi-atlas segmentation to separate different muscle groups and AT compartments. It has efficiently been used for body composition analysis of thousands of subjects in the UK Biobank imaging study. An example of segmentations of VAT, ASAT and different muscle groups using AMRA Researcher is shown in Figure 3.

A less automated generally available software tool that can be used for AT and muscle quantification is SliceOmatic (Tomovision, Quebec, Canada). This is a general interactive...
image segmentation tool that operates on one image slice at a time using a set of general-purpose image analysis tools to aid the operator. Even though a segmentation can be propagated from one slice to another, the time required for experienced operators was approximately 40 minutes for segmenting the complete VAT volume and 30 minutes for segmenting the calf muscles. Other similar tools are Analyze (AnalyzeDirect Inc. KS, USA), and Hippo Fat. Hippo Fat has been evaluated against SliceOMatic in the context of VAT segmentation. These three methods were also evaluated by Bonekamp et al.

Clinical Applications

Among the publications reviewed above, only seven had a specific clinical application in focus. Five were related to clinical aspects of abdominal AT distributions, one on intramuscular fat and one on both. Shen et al. investigated changes in volumes of abdominal organs and AT compartments in obese women during weight loss. Radmard et al. investigated the relationship between abdominal AT distribution and carotid atherosclerosis. Eichler et al. investigated lipodystrophy in HIV patients during anti-viral intervention. Karlsson et al. investigated gender differences in abdominal AT distribution in preschool children. Lareau-Trudel et al. measured intra-muscular fat in the legs in a study on facioscapulohumeral muscular dystrophy, and Orsso et al. investigated abdominal AT and intramuscular AT composition in youth with Prader-Willi syndrome.

In order to find more clinical applications of MRI-based quantification of AT, a wider search was made on PubMed for papers published in 2017 (as e-pub or print date) with "adipose tissue" and "MRI" in the title or abstract. Among the 137 papers found, 43 described clinical studies including quantification of VAT and/or SAT. Out of these 43 clinical studies, 27 used some kind of automated analysis, but most of those used tools with a low degree of automation. Hence, as many as 20 of these 27 studies used area measures in one or a few slices instead of measuring complete volumes of VAT and ASAT, likely due to the laborious task of using a 2D tool with low automation for analysing complete volumetric data. Among these area-based studies, SliceOMatic was the most common tool, used in 13 of the 43 clinical studies. Only a few studies used volumetric analysis tools.

The most common clinical areas in the studies found in this search were obesity (12), T2D (9), liver disease (8) and cardio-vascular disease (CVD) (5 studies). Six of the studies addressed BAT. The most common focus of the obesity studies concerned the effect of interventions (medical, surgical or life-style) on body composition. Most of the diabetes studies either investigated the effect of some intervention on abdominal AT distribution or the relationship between AT distribution and insulin resistance. The studies addressing liver fat either measured relations between body composition and liver fat or investigated intervention effects on liver fat and abdominal AT distribution. Most of the CVD-related studies investigated the relation between AT distribution and cardio-metabolic risk factors. The aims of the BAT studies were diverse. Two looked at hibernoma (a BAT tumour) and two looked at the effect of cold stimulation on BAT fat fraction.

Extending the search on PubMed a number of years back in time shows an approximately exponential increase in the number of publications with "adipose tissue" and "MRI" in title or abstract (see Figure 4). A search on NIH ClinicalTrials.gov on clinical trials with “adipose tissue” as outcome and “MRI” in other terms, shows a similar trend (see Figure 5). Both these search results indicate a rapidly growing interest for MRI in clinical studies related to adipose tissue.

One example of studies that would be extremely difficult, if indeed possible at all, to perform without highly automated analysis and quantitative MRI was recently published by Linge et al.
al. MR images from 6,021 subjects from the UK Biobank were analysed quantifying ASAT, VAT, thigh muscle volume, liver fat, and muscle fat infiltration and their multivariate associations to coronary heart disease and type 2 diabetes were investigated. The study showed that different diseases were linked to different imbalances in fat accumulation, which could not be described by sex, age, lifestyle, generalized adiposity or by investigating a single fat compartment alone. The method used in that study was AMRA® Researcher.

CONCLUSIONS

Apparently, up till now, rather few clinical studies have used highly automated methods for assessment of AT and muscle volume. Most of the studies used completely manual tools for segmentation. Some studies even used other imaging modalities (such as CT or DXA) to measure VAT even though an MRI scan was included in the study, thereby increasing the study time for the patient and also, in the case of CT or DXA, unnecessarily exposing the subjects to ionizing radiation.

Quantitative imaging both simplifies tissue segmentation and fat quantification. The fact that most studies in this review did not use quantitative imaging could to some extent be explained by a need to analyse images that had been previously acquired for other purposes. However, the frequent application of an intermediate inhomogeneity correction step to alleviate some of the problems with non-quantitative MRI, shows the importance of using quantitative MRI (e.g. FW-MRI) in studies involving body composition analysis.

One requirement for a method to be widely used in clinical studies is that it has been carefully evaluated, in particular with respect to repeatability and reproducibility. Another is that the method is readily accessible for other researchers and non-image analysis experts. A number of software packages for abdominal AT quantification have earlier been evaluated by Bonekamp et al., but all of them had a relatively low degree of automation. Still, the interest in using MRI in studies of adipose tissue is rapidly increasing, as can be seen in figures 4 and 5, in particular in studies related to obesity, T2D, liver disease and CVD.

As shown by Linge et al., simultaneous quantification of several body composition parameters reveals new relations between body fat distribution and different diseases. Such simultaneous volumetric analyses of multiple fat compartments and muscles would be a discouraging task, unless using automated analysis methods. The access to such methods enables efficient and detailed body composition assessment in patients with suspected metabolic disorders, particularly in cases when MRI is already an established examination, such as in liver diseases. In conclusion, there is a growing need for efficient and generally available methods for automated volumetric assessment of AT and muscle composition and, in particular, a need for reproducibility studies of such methods.

CONFLICTS OF INTEREST

The author is stockholder in AMRA Medical AB
REFERENCES


100. Fogel A, Goh AT, Fries LR, et al. Faster eating rates are associated with higher energy intakes during an Ad libitum meal, higher BMI and greater adiposity among 4.5 year old children – Results from the GUSTO cohort. The British journal of nutrition 2017; 117: 1042-1051. DOI: 10.1017/S0007114517000848.


FIGURES

Figure 1. Example of quantitative imaging. To the left is an original fat image acquired using 2-point Dixon fat-water separated imaging. To the right is the corresponding quantitative fat image after calibration using the fat-referenced method. Inhomogeneities in the intensity of adipose tissue can be observed in the original image (left) that are almost completely remove in the calibrated image (right).
Figure 2. Illustration of continuous and discrete quantification. Continuous methods (left column and black solid line in diagram) are less dependent on the image resolution than discrete methods (second column and dashed lines in diagram). See text for details.
Figure 3. Example of segmentation of adipose tissue and muscle groups using AMRA\textsuperscript{®} Researcher. To the left is the fat image with ASAT (blue) and VAT (red), and to the right is the water image with 10 different muscle groups coloured. Reproduced with permission from AMRA Medical AB.

Figure 4. Number of publications per year with "adipose tissue" and "MRI" in title or abstract found on PubMed.
Figure 5. Number of clinical trials per year with “adipose tissue” as outcome and “MRI” in other terms found on NIH ClinicalTrials.gov.
SEARCH METHOD

The search was made on Nov 13, 2017 on PubMed for publications containing "MRI" or "magnetic resonance imaging" and "automated", "automatic", "semiautomated" or "semiautomatic" and "fat" or "muscles" but not "brain" in the title or abstract using the following search string:

```
(((MRI[Title/Abstract] OR Magnetic resonance imaging[Title/Abstract]) AND (automated[Title/Abstract] OR automatic[Title/Abstract] OR semiautomated[Title/Abstract] OR semiautomatic[Title/Abstract]) AND (fat[Title/Abstract] OR muscle[Title/Abstract] OR muscles[Title/Abstract])) NOT brain[Title/Abstract])
```

Publications published before 2013 were excluded. From the remaining studies, 11 studies on breast MRI, two animal studies, one study on fetal imaging and one study on parapharyngeal fat pads were considered off-topic for this review and therefore excluded. The review by Hu et al.¹, which was discussed in the Review section, was removed from this summary since it discusses several different methods. Finally, 3 relevant studies²⁴ known to the author that did not appear in the PubMed search but published during that period were added to the list.

RESULTS

The search procedure described above resulted in 34 publications, summarized in the table below.
Table S1: Summary of publications 2013–2017 using automated analysis methods for MRI-based quantification of fat and muscles.

<table>
<thead>
<tr>
<th>Source</th>
<th>Measurements</th>
<th>Image pre-processing</th>
<th>Segmentation method</th>
<th>2D/3D analysis</th>
<th>Volumetric / single slice</th>
<th>Continuous / discrete</th>
<th>Automation</th>
<th>Validation</th>
<th>Repeatability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karlsson et al. 2013⁶</td>
<td>VAT, ASAT, TAAT, Total abdominal lean tissue</td>
<td>Inhomogeneity correction. Manual exclusion of liver.</td>
<td>Fuzzy C-means clustering to segment fat. Morphological operations to separate VAT from ASAT.</td>
<td>3D</td>
<td>Volume</td>
<td>Discrete</td>
<td>Semi-automated</td>
<td>Agreement with DXA⁹</td>
<td></td>
</tr>
<tr>
<td>Silver et al. 2013⁶</td>
<td>TAT, TAAT, TLST, VAT, ASAT</td>
<td>Threshholding and morphological operations</td>
<td></td>
<td>3D</td>
<td>Volume</td>
<td>Discrete</td>
<td>Fully</td>
<td>Agreement with DXA.</td>
<td></td>
</tr>
<tr>
<td>Thörmer et al. 2013⁷</td>
<td>VAT, ASAT</td>
<td>k-means clustering and 2D active contours to separate VAT from ASAT.</td>
<td>2D</td>
<td>Volume</td>
<td>Discrete</td>
<td>Fully or supervised</td>
<td>Repeatability for supervised method and accuracy against manual segmentation for fully automated and supervised method. CoV: 2.9% (VAT), 2.1% (ASAT); ICC: 0.997 (VAT and ASAT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valentinitsch et al. 2013²</td>
<td>Thigh muscles, inter-muscular AT and SAT area</td>
<td>k-means clustering and morphological operations</td>
<td></td>
<td>2D</td>
<td>Area</td>
<td>discrete</td>
<td>Fully automated</td>
<td>Comparison to manual segmentation</td>
<td></td>
</tr>
<tr>
<td>Thomas et al. 2014¹⁰</td>
<td>Total muscle volume excl. arms; left/right abdomen; left/right, posterior/anterior thigh; left/right lower leg.</td>
<td>Fat-referenced image calibration</td>
<td>3D multi-atlas segmentation based on non-rigid registration</td>
<td>3D</td>
<td>Volume</td>
<td>Continuous</td>
<td>Fully automated</td>
<td>Comparison to manual thresholding and test-retest repeatability</td>
<td>ICC: 0.99 - 1.0 for all compartments. 95% LoA: 0.32 - 0.20 L for total muscle volume</td>
</tr>
<tr>
<td>Wang et al. 2015¹¹</td>
<td>VAT, ASAT</td>
<td>Inhomogeneity correction</td>
<td>K-means clustering and 2D active contours.</td>
<td>2D</td>
<td>Volume</td>
<td>Discrete</td>
<td>Fully</td>
<td>Comparison with manually set thresholds.</td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Org &amp; Segmentation</td>
<td>Preparation</td>
<td>Labeling</td>
<td>Dimensions</td>
<td>Segmentation</td>
<td>Validation</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------</td>
<td>-------------</td>
<td>----------</td>
<td>------------</td>
<td>--------------</td>
<td>------------</td>
<td>-----------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eichler et al. 2015</td>
<td>VAT, ASAT, TAAT</td>
<td>None</td>
<td>Manual thresholding and 2D manual delineation of VAT</td>
<td>2D Area Discrete Semi-automated</td>
<td>ROC analysis of ability to separate patient groups.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addeman et al. 2015</td>
<td>VAT, ASAT, TAAT</td>
<td>Removal of background and bone marrow</td>
<td>Thresholding of FF images, 3D segmentation of abdominal muscles.</td>
<td>3D Volume Continuous Fully</td>
<td>CoV for within-day repeatability and one-week reproducibility. Bland-Altman analysis of agreement with manual segmentation. Data from three different cohorts.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borga et al. 2015</td>
<td>VAT, ASAT, TAAT</td>
<td>Fat-referenced image calibration</td>
<td>3D multi-atlas segmentation based on non-rigid registration.</td>
<td>3D Volume Continuous Supervised Agreement with manual segmentation and inter- and intra-observer repeatability.</td>
<td>Intra-observer CoV: 1.6% (VAT), 1.1% (ASAT), Inter-observer CoV: 1.4% (VAT), 1.2% (ASAT). Inter-observer ICC 0.999 (VAT and ASAT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lareau-Trudel et al. 2015</td>
<td>Intra-muscular fat fraction in legs</td>
<td>k-means clustering, morphological operations and active contours</td>
<td>2D Area discrete Supervised Comparison to visual IMAT scoring and inter- and intra-op variability for manual correction</td>
<td>ICC: 0.97 (inter-observer), 0.98 (intra-observer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andrews et al. 2015</td>
<td>11 individual muscles in the thigh</td>
<td>Statistical shape model</td>
<td>2D Volume discrete Fully automated Comparison to another automated segmentation method.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karlsson et al. 2015</td>
<td>Lean muscle tissue volume of 10 different muscle L/R posterior/anterior thigh, L/R abdomen, L/R arm</td>
<td>Fat-referenced image calibration</td>
<td>3D Volume Continuous Fully automated Comparison to manual segmentation and reproducibility between 1.5T and 3T scanners</td>
<td>ICC: 0.99. LoA: -0.86 - 2.39 L (all muscles, between scanners)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orgiu et al. 2015</td>
<td>Thigh muscles, inter-muscular AT and SAT</td>
<td>Inhomogeneity correction Fuzzy c-means clustering and active contours</td>
<td>2D Area discrete Fully automated Comparison to manual segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun et al.</td>
<td>VAT, ASAT, TAAT</td>
<td>Inhomogeneity Fuzzy C-means clustering to segment fat and 2D</td>
<td>2D Volume Discrete Fully Bland-Altman and linear correlation against manual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Type</td>
<td>Methodology</td>
<td>Volume</td>
<td>Discrete/Continuous</td>
<td>Automation</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------</td>
<td>---------------------</td>
<td>------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>VAT, ASAT</td>
<td>Inhomogeneity correction Fuzzy C-means clustering to segment fat and 2D active contours to separate VAT from ASAT.</td>
<td>2D</td>
<td>Area</td>
<td>Semi-automated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gifford et al. 2016</td>
<td>FF and R2* calculations PET detection of BAT transferred to co-registered MR images</td>
<td>3D</td>
<td>Volume</td>
<td>Semi-automated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mhuiris et al. 2016</td>
<td>AT infiltration in 4 sections of the lumbar paravertebral muscles Manual segmentation of muscle, automated subdivision in 4 sections.)</td>
<td>2D</td>
<td>Area</td>
<td>Semi-automated</td>
<td>Intra- and inter-observer repeatability of muscle fat infiltration ICC: 0.88 (intra-observer), 0.82 (inter-observer). LoA: -5.48 - 4.92 % (intra-observer), -6.85 - 5.90 % (inter-observer).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yang et al. 2016</td>
<td>Intermuscular AT, SAT and muscle volume of thigh muscles Inhomogeneity correction Machine learning classifier and morphological operations</td>
<td>3D</td>
<td>Volume</td>
<td>Fully automated</td>
<td>Comparison to manual segmentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Le Troter et al. 2016</td>
<td>Intermuscular AT, SAT and muscle volume of quadriceps femoris 3D multi-atlas segmentation based on non-rigid registration</td>
<td>3D</td>
<td>Volume</td>
<td>fully and semi-automated</td>
<td>Comparison to manual segmentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ugart et al. 2016</td>
<td>Intramuscular AT and connective tissue in calf muscles Inhomogeneity correction and manual removal of SAT Fuzzy c-means clustering of feature vectors computed from dual echo and structure tensor information</td>
<td>3D</td>
<td>Volume</td>
<td>semi-automated (requires manual pre-processing)</td>
<td>Accuracy evaluated against digital phantom and manual segmentation of in-vivo data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fortin et al. 2017</td>
<td>Paraspinal muscle and AT area Thresholding</td>
<td>2D</td>
<td>Area</td>
<td>semi-automated (Manual definition of muscle ROI9) Comparison to manual thresholding and intra-observer repeatability ICC: 0.78 - 0.97 (IMAT), 0.84-0.94 (lean muscle), SEM: 0.41 - 0.70 (IMAT), 0.49 - 0.67 (lean muscle)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lundström et al. 2017</td>
<td>BAT FF and R2* calculations 3D multi-atlas segmentation based on non-rigid registration</td>
<td>3D</td>
<td>Volume</td>
<td>Fully</td>
<td>Dice coefficient and comparison of FF and R2* with a semi-automated method.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maddalao et al. 2017</td>
<td>VAT, ASAT Thresholding based on grey-scale clustering 2D morphology</td>
<td>2D</td>
<td>Volume</td>
<td>Fully</td>
<td>Bland-Altman, linear correlation and volume difference % against manual segmentation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Target Tissue</td>
<td>Methodology</td>
<td>2D/3D</td>
<td>Volume</td>
<td>Boundary</td>
<td>Variability</td>
<td>Comparison</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>-------------</td>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulbrich et al. 2017&lt;sup&gt;30&lt;/sup&gt;</td>
<td>TAT, VAT, ASAT, TAAT, TLMT&lt;sup&gt;8&lt;/sup&gt; muscle groups: left/right abdomen; left/right, posterior/anterior thigh; left/right lower leg.</td>
<td>Fat-referenced image calibration</td>
<td>3D</td>
<td>multi-atlas segmentation based on non-rigid registration.</td>
<td>3D</td>
<td>Volume</td>
<td>Continuous</td>
<td>Supervised</td>
<td>Comparison to anthropometric measures and bio-impedance analysis.</td>
</tr>
<tr>
<td>Kim et al. 2017&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Supraspinatus volume</td>
<td>Active contours</td>
<td>2D</td>
<td>Volume</td>
<td>semi-automated (requires manual initialization)</td>
<td></td>
<td></td>
<td></td>
<td>Comparison to manual segmentation on 5 subjects</td>
</tr>
<tr>
<td>Waduud et al. 2017&lt;sup&gt;32&lt;/sup&gt;</td>
<td>VAT, ASAT, TAAT</td>
<td>2D thresholding and manual segmentation of VAT and ASAT.</td>
<td>2D</td>
<td>Area</td>
<td>Discrete</td>
<td>Fully</td>
<td></td>
<td>Bland-Altman and linear correlation of intra- and inter-observer variability.</td>
<td></td>
</tr>
<tr>
<td>Middleton et al. 2017&lt;sup&gt;4&lt;/sup&gt;</td>
<td>VAT, ASAT, TAAT, liver fat fraction and thigh muscle volume</td>
<td>Fat-referenced image calibration</td>
<td>3D</td>
<td>multi-atlas segmentation based on non-rigid registration.</td>
<td>3D</td>
<td>Volume</td>
<td>Continuous</td>
<td>Supervised</td>
<td>ICC and CoV for test-retest repeatability.</td>
</tr>
<tr>
<td>Orsso et al. 2017&lt;sup&gt;33&lt;/sup&gt;</td>
<td>VAT, ASAT, IMAT, abdominal muscles</td>
<td>2D morphology</td>
<td>2D</td>
<td>Volume</td>
<td>Discrete</td>
<td>Semi-automated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kemnitz et al. 2017&lt;sup&gt;34&lt;/sup&gt;</td>
<td>Thigh muscles, IMAT&lt;sup&gt;2&lt;/sup&gt; and SAT area</td>
<td>Histogram equalization</td>
<td>Active contours</td>
<td>2D</td>
<td>Area</td>
<td>discrete</td>
<td>Supervised</td>
<td>Comparison to manual segmentation (20 subjects) and inter-observer repeatability</td>
<td>CoV 0.9% (SAT), 1.9% (IMAT), 0.4% - 0.9% (muscles)</td>
</tr>
<tr>
<td>Hui et al. 2017&lt;sup&gt;35&lt;/sup&gt;</td>
<td>VAT, ASAT, TAAT</td>
<td>Intensity correction using N4ITK</td>
<td>2D</td>
<td>Volume</td>
<td>Discrete</td>
<td>Fully</td>
<td>Bland-Altman and ICC against manual segmentation.</td>
<td>ICC: VAT: 0.984; ASAT: 0.997</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


