A comparative study of software programmes for cross-sectional skeletal muscle and adipose tissue measurements on abdominal computed tomography scans of rectal cancer patients

van Vugt, Jeroen L.A.; Levolger, Stef; Gharbharan, Arvind; Koek, Marcel; Niessen, Wiro J.; Burger, Jacobus W.A.; Willemsen, Sten P.; de Bruin, Ron W.F.; IJzermans, Jan N.M.

DOI
10.1002/jcsm.12158

Publication date
2017

Document Version
Final published version

Published in
Journal of Cachexia, Sarcopenia and Muscle

Citation (APA)

Important note
To cite this publication, please use the final published version (if applicable). Please check the document version above.
A comparative study of software programmes for cross-sectional skeletal muscle and adipose tissue measurements on abdominal computed tomography scans of rectal cancer patients

Jeroen L.A. van Vugt1*, Stef Levolger1, Arvind Gharbharan1, Marcel Koek2, Wiro J. Niessen2,3, Jacobus W.A. Burger4, Sten P. Willemsen5, Ron W.F. de Bruin1 & Jan N.M. IJzermans1*

1Department of Surgery, Erasmus MC University Medical Centre, Rotterdam, the Netherlands; 2Department of Radiology and Medical Informatics, Erasmus MC University Medical Centre, Rotterdam, the Netherlands; 3Faculty of Applied Sciences, Delft University of Technology, Delft, the Netherlands; 4Department of Surgical Oncology, Erasmus MC – Daniel den Hoed Cancer Institute, Rotterdam, the Netherlands; 5Department of Biostatistics, Erasmus MC University Medical Centre, Rotterdam, the Netherlands

Abstract

Background  The association between body composition (e.g. sarcopenia or visceral obesity) and treatment outcomes, such as survival, using single-slice computed tomography (CT)-based measurements has recently been studied in various patient groups. These studies have been conducted with different software programmes, each with their specific characteristics, of which the inter-observer, intra-observer, and inter-software correlation are unknown. Therefore, a comparative study was performed.

Methods  Fifty abdominal CT scans were randomly selected from 50 different patients and independently assessed by two observers. Cross-sectional muscle area (CSMA, i.e. rectus abdominis, oblique and transverse abdominal muscles, paraspinal muscles, and the psoas muscle), visceral adipose tissue area (VAT), and subcutaneous adipose tissue area (SAT) were segmented by using standard Hounsfield unit ranges and computed for regions of interest. The inter-software, intra-observer, and inter-observer agreement for CSMA, VAT, and SAT measurements using FatSeg, OsiriX, ImageJ, and sliceOmatic were calculated using intra-class correlation coefficients (ICCs) and Bland–Altman analyses. Cohen’s κ was calculated for the agreement of sarcopenia and visceral obesity assessment. The Jaccard similarity coefficient was used to compare the similarity and diversity of measurements.

Results  Bland–Altman analyses and ICC indicated that the CSMA, VAT, and SAT measurements between the different software programmes were highly comparable (ICC 0.979–1.000, P < 0.001). All programmes adequately distinguished between the presence or absence of sarcopenia (κ = 0.88–0.96 for one observer and all κ = 1.00 for all comparisons of the other observer) and visceral obesity (all κ = 1.00). Furthermore, excellent intra-observer (ICC 0.999–1.000, P < 0.001) and inter-observer (ICC 0.998–0.999, P < 0.001) agreement for all software programmes were found. Accordingly, excellent Jaccard similarity coefficients were found for all comparisons (mean ≥ 0.964).

Conclusions  FatSeg, OsiriX, ImageJ, and sliceOmatic showed an excellent agreement for CSMA, VAT, and SAT measurements on abdominal CT scans. Furthermore, excellent inter-observer and intra-observer agreement were achieved. Therefore, results of studies using these different software programmes can reliably be compared.

Keywords  Body composition; Skeletal muscle mass; Adipose tissue mass; Comparability; Software; Computed tomography

© 2016 The Authors. Journal of Cachexia, Sarcopenia and Muscle published by John Wiley & Sons Ltd on behalf of the Society on Sarcopenia, Cachexia and Wasting Disorders

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
Introduction

Biological frailty and analytic morphomics (i.e. body composition) have increasingly gained interest in recent years in relation to treatment outcomes, such as complications and (disease-free) survival.1,2 Frailty, a state of increased vulnerability towards stressors, leads to an increased risk of developing adverse health outcomes3 and is an important predictor of complications after interventional procedures, such as surgery and chemotherapy.4–7 For example, frail patients undergoing colorectal surgery have a fourfold increased risk to develop major post operative complications.8 One of the hallmark signs of frailty is sarcopenia, the involuntary depletion of skeletal muscle mass.8–11 It is estimated that up to 25% of persons under 70 years of age and over 50% of persons of 80 years and older experience sarcopenia.12 In addition, up to 80% of patients with advanced cancer are affected by cancer-induced cachexia, a clinical condition that also results in skeletal muscle wasting with or without the loss of body fat.13–15 Patients with cachexia are more prone to a reduced therapy effect,16 and patients with low skeletal muscle mass experience increased chemotherapy toxicity.17,18 This ultimately results in death in nearly one-third of all cancer patients.19–22

Over the last years, numerous studies have used abdominal computed tomography (CT) scans to quantify skeletal muscle mass, for example, in clinical17,18,23–25 and surgical26 oncology, vascular surgery,27 and transplantation surgery28,29 patients. Furthermore, multiple studies measured visceral and/or subcutaneous adipose tissue on CT scans.30–33 However, different software programmes have been used to perform these body composition analyses, such as FatSeg,33 OsiriX,7 ImageJ,24 and sliceOmatic.23 To be able to adequately compare study results, the comparability of these various software programmes should be known. Therefore, the aim of this study was to investigate the agreement of these four different software packages for the assessment of cross-sectional skeletal muscle and subcutaneous and visceral adipose tissue measurements on abdominal CT scans.

Materials and Methods

Patients

Fifty abdominal CT scans of patients who were scheduled for rectal cancer resection at Erasmus MC University Medical Centre (Rotterdam, the Netherlands) between 2005 and 2012 were randomly selected. All CT scans were routinely performed as part of the pre-operative diagnostic work up or assessment of down staging after neo-adjuvant therapy. Only one CT scan was used per patient. None of the patients had an ostomy, abdominal wall deformity, abdominal wall tumour, or a CT scan with artefacts at the level of L3 that could potentially influence measurements. Self-reported weight and height in the pre-operative workup were retrospectively collected from electronic patient files.

Skeletal muscle and adipose tissue area measurements

The cross-sectional skeletal muscle area (CSMA), subcutaneous adipose tissue area (SAT), and visceral adipose tissue area (VAT) (cm²), including renal adipose tissue, were measured at the mid third lumbar vertebra (L3) level on a slice showing both transversal processes. The CSMA measurements included the following muscles: psoas, paraspinal, transverse abdominal, external oblique, internal oblique, and rectus abdominis. All abdominal CT scans were assessed on identical slices in a random order by two medically trained observers [AG (observer A) and JLAvV (observer B)], with great knowledge about radiological anatomy and extensive experience in skeletal muscle and adipose tissue area measurements using various software programmes. Observer A performed measurements twice on identical a priori selected slices, whereas observer B performed a second reading without a priori selected slice numbers. The observers were blinded for each other’s measurements and for patient details. For each observer, the time interval between two readings in the same patient with different software programmes was at least 1 week. This resulted in an interval of at least 4 weeks between two readings within one patient with the same software programme. Only the first reading of observer B was used for the inter-software and inter-observer comparisons.

The CSMA was corrected for height squared (m²), resulting in the L3 muscle index (SMI, cm²/m²). Patients were classified as having sarcopenia or not having sarcopenia according to previously described cut-off values (52.4 cm²/m² for men and 38.5 cm²/m² for women).23 Predefined cut-off values for VAT to define visceral obesity of 163.8 cm² for men and 80.1 cm² for women were used.24 For subcutaneous adipose tissue, no cut-off values have been reported in the literature.

Four software programmes were compared: FatSeg [developed by the Biomedical Imaging Group Rotterdam of Erasmus MC, Rotterdam, the Netherlands, using MeVisLab (Mevis Medical Solutions, Bremen, Germany)], OsiriX (Pixmeo SARL, Geneva, Switzerland), ImageJ (National Institutes of Health, Bethesda, MD, USA), and sliceOmatic (TomoVision, Magog, Canada). The CSMA, VAT, and SAT were segmented using standard Hounsfield unit (HU) thresholds in all four software programmes. An intensity window between −30 and +150 HU was used for skeletal muscle tissue.35 For adipose tissue, an intensity window between −190 and −30 HU was used.36 Because the tissue of interest is manually selected, competency in anatomic radiology is a prerequisite for these measurements.
**FatSeg**

FatSeg is an in-house developed software programme to perform soft tissue measurements on CT scans and was developed using the MeVisLab development environment for medical image processing and visualization version 2.4 (available from http://www.mevislab.de). Inner and outer contours of aforementioned skeletal muscle and adipose tissue regions were manually traced. The skeletal muscle and adipose tissue areas were computed automatically using the preset HU intensity thresholds and expressed in square centimetre. Intraluminal contents initially marked as adipose tissue were manually erased. Cutaneous tissue was included in the SAT measurement. Measurements were performed on a 3.2 GHz Intel® Core™ i5 MacBook Air (Apple Inc., Cupertino, CA, USA) personal computer.

**OsiriX**

The open-source 32-bit edition of OsiriX version 5.8.5 (available from http://www.osirix-viewer.com) was used. The ‘Grow Region (2D/3D Segmentation)’ tool was used to semi-automatically select skeletal muscle and adipose tissue regions within our preset HU intensity thresholds. Non-skeletal muscle tissue regions adjacent to skeletal muscle were manually removed from the area selection using the brush option. The brush option was also used to manually erase intraluminal areas with contents having radiological density between −190 and −30 HU, resembling fatty content. Cutaneous tissue was not included in the SAT measurement. The skeletal muscle and adipose tissue areas were computed automatically and expressed in square centimetre using a 1.3 GHz Intel® Core™ i5 MacBook Air (Apple Inc., Cupertino, CA, USA) and computer mouse.

**ImageJ**

ImageJ version 1.48 is a freely downloadable public domain software programme developed by the National Institutes of Health for image processing and analyzing (available from http://rsbweb.nih.gov/ij/download.html). First, manual delineation of the outer contour of the abdominal wall and paraspinal muscles was performed, and the surface area of tissue with an attenuation between −30 and +150 HU was computed automatically (mm²) and manually divided by 100, resulting in square centimetre. Second, delineation of the inner contour of the abdominal wall, paraspinal, and psoas muscles was performed in a similar fashion to allow for subsequent correction of intra-abdominal content with attenuation between the preset HU intensity thresholds. The inner contour was manually subtracted from the outer contour surface area, resulting in the cross-sectional skeletal muscle area (cm²).37 The SAT measurements were performed in a similar manner as the muscle measurements, whereas VAT measurements were performed by delineating a contour through the inner contour of the abdominal wall muscles, psoas muscles, and vertebrae followed by manual erasing of intraluminal fatty content. Cutaneous tissue was included in the SAT measurement. A 3.2 GHz Intel® Core™ i5 Dell personal computer was used.

**SliceOmatic**

SliceOmatic (Tomovision, Magog, Canada) version 5.0 (64 bit; available from http://www.tomovision.com) was used. Tissue was semi-automatically selected with the ‘Region Growing’ mode using the ‘Grow 2D’ and ‘Paint’ tools. Non-skeletal muscle tissue regions adjacent to skeletal muscle having radiological density between the predefined HU thresholds were manually erased using the ‘Paint’ tool. Cutaneous tissue was included in the SAT measurement. A 3.2 GHz Intel® Core™ i5 Dell personal computer was used.

**Cutaneous tissue disclosure**

In OsiriX, cutaneous tissue is not included in the SAT measurement, because this is not automatically selected using the ‘Grow Region (2D/3D Segmentation)’ tool. SliceOmatic also allows to exclude encompassed skin. However, not all software programmes allow to reliably exclude cutaneous tissue from SAT as a consequence of their measurement method: the delineation of tissue of interest using inner and outer contours. Consequently, to ensure highly comparable measurements in three rather than two software programmes, cutaneous tissue was included in the SAT measurements with sliceOmatic. A comparison of SAT measurements using sliceOmatic with and without the inclusion of cutaneous tissue resulted in a median difference of 2.3% (interquartile range 0.8–3.8) and was considered acceptable.

**Statistical analysis**

Continuous data are presented as mean with the standard error of the mean. Normality was tested using the Shapiro–Wilk test. Differences between the different software packages and within and between observers were compared using the paired samples t-test for normally distributed data and the Wilcoxon signed rank test for data that were not normally distributed. The inter-software and inter-observer correlation for the cross-sectional skeletal muscle, visceral adipose tissue, and subcutaneous adipose tissue measurements were calculated using intra-class correlation coefficients (ICCs) with 95% confidence interval using a two-way mixed single measures model with absolute agreement. For the inter-observer correlation, reading 1 of...
observer B was compared with reading 1 of observer A. Ninety-five per cent limits of agreement were determined to investigate the agreement between the various software programmes, according to the method described by Bland and Altman. The presence of proportional systematic bias was determined by linear regression analysis of the difference and mean of two measurements. The inter-software, inter-observer and intra-observer agreement of the assessment of sarcopenia and visceral obesity were analyzed using Cohen’s κ coefficients. The ICCs and Cohen’s κ coefficients were interpreted as poor (0.00–0.49), fair to good (0.50–0.74), and excellent (0.75–1.00), as proposed by Shrout and Fleiss. The Jaccard similarity coefficient, ranging from 0 to 1, was used to compare the similarity and diversity of measurements by dividing the area of the intersection by the size of the union of two measurements. An overlay of two measurements was created, and the Jaccard similarity coefficient was calculated using MeVisLab version 2.7.1 (MeVis Medical Solutions AG, Bremen, Germany). A Jaccard similarity coefficient of 1 represents perfect overlap of two samples, whereas 0 represents no overlap. Two-tailed P-values <0.05 were considered statistically significant. All statistical analyses were performed using IBM SPSS Statistics for Windows version 21.0 (IBM Corp. Armonk, NY, USA).

Results

Patients

The study population consisted of 29 men (58%) and 21 women (42%) with a median age of 62 years (range 33–81) and a median body mass index of 24.6 kg/m² (range 16.5–38.8). Ten patients had stage II (20.0%), 24 stage III (48.0%), and 15 stage IV (30.0%) rectal cancer. Tumour stage was unknown for one patient. The mean CSMA, VAT, and SAT for all measurements are provided in Tables 1 and 2.

Inter-software agreement

The inter-software ICCs were excellent (≥0.999) for the CSMA, VAT, and SAT for all software programmes with P-values <0.001 (Table 3). Figure 1 and Supporting Information, Figures S1 and S2 show the Bland–Altman 95% limits of agreement plots, with the mean difference and 95% limits of agreement for the CSMA, VAT, and SAT for both observers. All plots show a good agreement between the various software programmes. Small limits of agreement are observed in the CSMA measurements, whereas these limits of agreement are greater for the VAT and SAT measurements. Proportional systematic bias was observed between FatSeg and OsiriX for CSMA (P=0.049) for observer B (Figure 1A) and between FatSeg and sliceOmatic for SAT (P=0.031) for observer A (Supporting Information, Figure S2C). Furthermore, proportional systematic bias was frequently observed between programmes for VAT measurements (Supporting Information, Figure S1). Comparable results were achieved when non a priori selected slices of observer B (reading 2) were analyzed (data not shown). The mean Jaccard similarity coefficients for the inter-software comparisons are summarized in Table 4 and depicted in Figure 2.

### Table 1

<table>
<thead>
<tr>
<th>Software</th>
<th>Reading 1 (cm²)</th>
<th>SEM</th>
<th>Reading 2 (cm²)</th>
<th>SEM</th>
<th>Mean difference (95% CI)</th>
<th>P-value</th>
<th>ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observer A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Skeletal muscle area</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FatSeg</td>
<td>139.0</td>
<td>5.2</td>
<td>139.3</td>
<td>5.2</td>
<td>−0.3 (−0.6; 0.0)</td>
<td>0.072b</td>
<td>0.999 (0.999–1.000)</td>
</tr>
<tr>
<td>OsiriX</td>
<td>139.4</td>
<td>5.1</td>
<td>138.7</td>
<td>5.1</td>
<td>0.7 (0.4; 1.0)</td>
<td>&lt;0.001b</td>
<td>0.999 (0.999–1.000)</td>
</tr>
<tr>
<td>ImageJ</td>
<td>139.0</td>
<td>5.2</td>
<td>139.3</td>
<td>5.1</td>
<td>−0.3 (−0.6; −0.1)</td>
<td>0.013b</td>
<td>1.000 (0.999–1.000)</td>
</tr>
<tr>
<td>sliceOmatic</td>
<td>138.7</td>
<td>5.2</td>
<td>138.6</td>
<td>5.2</td>
<td>0.1 (−0.2; 0.4)</td>
<td>0.441b</td>
<td>1.000 (0.999–1.000)</td>
</tr>
<tr>
<td><strong>Visceral adipose tissue area</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FatSeg</td>
<td>149.9</td>
<td>13.1</td>
<td>149.2</td>
<td>13.1</td>
<td>0.7 (0.3; 1.0)</td>
<td>&lt;0.001b</td>
<td>1.000 (1.000–1.000)</td>
</tr>
<tr>
<td>OsiriX</td>
<td>147.6</td>
<td>13.0</td>
<td>147.3</td>
<td>13.0</td>
<td>0.3 (−0.3; 0.8)</td>
<td>0.220b</td>
<td>1.000 (1.000–1.000)</td>
</tr>
<tr>
<td>ImageJ</td>
<td>148.6</td>
<td>13.0</td>
<td>150.8</td>
<td>12.8</td>
<td>−2.2 (−7.5; 3.1)</td>
<td>0.003b</td>
<td>0.979 (0.964–0.988)</td>
</tr>
<tr>
<td>sliceOmatic</td>
<td>147.1</td>
<td>13.0</td>
<td>146.6</td>
<td>13.0</td>
<td>0.5 (0.2; 0.9)</td>
<td>0.004b</td>
<td>1.000 (1.000–1.000)</td>
</tr>
<tr>
<td><strong>Subcutaneous adipose tissue area</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FatSeg</td>
<td>158.9</td>
<td>11.2</td>
<td>158.9</td>
<td>11.2</td>
<td>0.1 (−0.2; 0.3)</td>
<td>0.359b</td>
<td>1.000 (1.000–1.000)</td>
</tr>
<tr>
<td>OsiriX</td>
<td>155.9</td>
<td>11.2</td>
<td>155.7</td>
<td>11.3</td>
<td>0.2 (−0.1; 0.4)</td>
<td>0.137b</td>
<td>1.000 (1.000–1.000)</td>
</tr>
<tr>
<td>ImageJ</td>
<td>158.9</td>
<td>11.2</td>
<td>159.1</td>
<td>11.3</td>
<td>−0.2 (−0.5; 0.0)</td>
<td>0.201b</td>
<td>1.000 (1.000–1.000)</td>
</tr>
<tr>
<td>sliceOmatic</td>
<td>158.8</td>
<td>11.3</td>
<td>158.8</td>
<td>11.3</td>
<td>0.0 (−0.3; 0.2)</td>
<td>0.448b</td>
<td>1.000 (1.000–1.000)</td>
</tr>
</tbody>
</table>

CI confidence interval; ICC, intra-class correlation coefficient; SEM, standard error of the mean.

bCalculated with Wilcoxon signed rank test.

DOI: 10.1002/jcsm.12158
Table 2  Mean cross-sectional skeletal muscle and visceral and subcutaneous adipose tissue area (cm²) measurements and inter-observer agreement indices (i.e. ICC) using FatSeg, OsiriX, ImageJ, and sliceOmatic of reading 1 of observers A and B

<table>
<thead>
<tr>
<th>Software program</th>
<th>Observer A</th>
<th>Observer B</th>
<th>Mean difference (95% CI)</th>
<th>P-value</th>
<th>ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FatSeg (cm²)</td>
<td>Reading 1</td>
<td>Reading 1</td>
<td>Mean difference (cm²)</td>
<td>P-value</td>
<td>ICC (95% CI)</td>
</tr>
<tr>
<td>FatSeg</td>
<td>149.9</td>
<td>148.7</td>
<td>1.2 (0.8; 1.5)</td>
<td>&lt;0.001</td>
<td>1.000 (0.999-1.000)</td>
</tr>
<tr>
<td>OsiriX</td>
<td>147.6</td>
<td>147.3</td>
<td>0.3 (0.3; 0.8)</td>
<td>0.133</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>ImageJ</td>
<td>148.6</td>
<td>148.4</td>
<td>0.3 (0.1; 0.6)</td>
<td>0.015</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>sliceOmatic</td>
<td>147.1</td>
<td>146.9</td>
<td>0.2 (0.1; 0.5)</td>
<td>0.412</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>Visceral adipose tissue area (VAT)</td>
<td>Reading 1</td>
<td>Reading 1</td>
<td>Mean difference (cm²)</td>
<td>P-value</td>
<td>ICC (95% CI)</td>
</tr>
<tr>
<td>FatSeg</td>
<td>158.9</td>
<td>159.2</td>
<td>–0.3 (–0.5; –0.1)</td>
<td>0.005</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>OsiriX</td>
<td>155.9</td>
<td>155.8</td>
<td>0.1 (–0.3; 0.5)</td>
<td>0.918</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>ImageJ</td>
<td>158.9</td>
<td>158.7</td>
<td>0.2 (–0.2; 0.5)</td>
<td>0.306</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>sliceOmatic</td>
<td>158.8</td>
<td>158.5</td>
<td>0.2 (0.0; 0.5)</td>
<td>0.183</td>
<td>1.000 (1.000-1.000)</td>
</tr>
</tbody>
</table>

Intra-observer and inter-observer agreement

The ICCs for the intra-observer agreement of observer A were all 0.979 or higher for the different software programmes, approaching perfect correlation (Table 1). The ICCs for the inter-observer agreement also approached perfect agreement (all 0.999, Table 2). The mean CSMA was significantly lower for observer A compared with observer B for all software programmes. A significantly higher mean VAT of observer A was found using FatSeg (149.9 cm² vs. 148.7 cm², P < 0.001) and ImageJ (148.6 cm² vs. 148.4 cm², P = 0.015) compared with observer B, whereas the mean VAT of both observers did not significantly differ for OsiriX (P = 0.133) and sliceOmatic (P = 0.412). The mean SAT did significantly differ for FatSeg (158.9 cm² vs. 159.2 cm², P = 0.005) between the observers. Comparable results were observed when non a priori selected slices of reading 2 of observer B were used for analyses (data not shown). The mean Jaccard

Table 3  Mean cross-sectional skeletal muscle and visceral and subcutaneous adipose tissue area (cm²) measurements and inter-software agreement indices (i.e. ICC) using FatSeg, OsiriX, ImageJ, and sliceOmatic of reading 1 of observer B

<table>
<thead>
<tr>
<th>Software programme</th>
<th>Mean difference (cm²) (95% CI)</th>
<th>SEM</th>
<th>P-value</th>
<th>ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skeletal muscle area (CSMA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FatSeg–OsiriX</td>
<td>–0.4 (–0.8; 0.0)</td>
<td>0.184</td>
<td>0.047</td>
<td>0.999 (0.999-1.000)</td>
</tr>
<tr>
<td>FatSeg–ImageJ</td>
<td>0.0 (–0.3; 0.3)</td>
<td>0.151</td>
<td>0.992</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>FatSeg–sliceOmatic</td>
<td>0.3 (–0.2; 0.8)</td>
<td>0.230</td>
<td>0.207</td>
<td>0.999 (0.998-0.999)</td>
</tr>
<tr>
<td>OsiriX–ImageJ</td>
<td>0.4 (0.1; 0.7)</td>
<td>0.161</td>
<td>0.023</td>
<td>0.999 (0.999-1.000)</td>
</tr>
<tr>
<td>OsiriX–sliceOmatic</td>
<td>0.7 (0.3; 1.1)</td>
<td>0.189</td>
<td>0.001</td>
<td>0.999 (0.998-1.000)</td>
</tr>
<tr>
<td>ImageJ–sliceOmatic</td>
<td>0.3 (–0.1; 0.7)</td>
<td>0.208</td>
<td>0.165</td>
<td>0.999 (0.999-1.000)</td>
</tr>
<tr>
<td>Visceral adipose tissue area (VAT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FatSeg–OsiriX</td>
<td>2.3 (1.6; 2.9)</td>
<td>0.326</td>
<td>&lt;0.001</td>
<td>0.999 (0.995-1.000)</td>
</tr>
<tr>
<td>FatSeg–ImageJ</td>
<td>1.2 (0.8; 1.7)</td>
<td>0.203</td>
<td>&lt;0.001</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>FatSeg–sliceOmatic</td>
<td>2.8 (2.3; 3.2)</td>
<td>0.238</td>
<td>&lt;0.001</td>
<td>0.999 (0.971-1.000)</td>
</tr>
<tr>
<td>OsiriX–ImageJ</td>
<td>–1.0 (–1.5; –0.6)</td>
<td>0.237</td>
<td>&lt;0.001</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>OsiriX–sliceOmatic</td>
<td>0.5 (0.0; 0.9)</td>
<td>0.229</td>
<td>0.044</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>ImageJ–sliceOmatic</td>
<td>1.5 (1.2; 1.8)</td>
<td>0.158</td>
<td>&lt;0.001</td>
<td>1.000 (0.995-1.000)</td>
</tr>
<tr>
<td>Subcutaneous adipose tissue area (SAT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FatSeg–OsiriX</td>
<td>3.0 (2.5; 3.6)</td>
<td>0.256</td>
<td>&lt;0.001</td>
<td>0.999 (0.948-1.000)</td>
</tr>
<tr>
<td>FatSeg–ImageJ</td>
<td>0.1 (–0.3; 0.4)</td>
<td>0.180</td>
<td>0.698</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>FatSeg–sliceOmatic</td>
<td>0.2 (–0.1; 0.5)</td>
<td>0.141</td>
<td>0.240</td>
<td>1.000 (1.000-1.000)</td>
</tr>
<tr>
<td>OsiriX–ImageJ</td>
<td>–3.0 (–3.5; –2.5)</td>
<td>0.260</td>
<td>&lt;0.001</td>
<td>0.999 (0.956-1.000)</td>
</tr>
<tr>
<td>OsiriX–sliceOmatic</td>
<td>–2.9 (–3.3; –2.5)</td>
<td>0.211</td>
<td>&lt;0.001</td>
<td>0.999 (0.932-1.000)</td>
</tr>
<tr>
<td>ImageJ–sliceOmatic</td>
<td>0.1 (–0.2; 0.4)</td>
<td>0.139</td>
<td>0.485</td>
<td>1.000 (1.000-1.000)</td>
</tr>
</tbody>
</table>

The results of observer A are comparable with those of observer B.
CI confidence interval; ICC, inter-class and intra-class correlation coefficients; SEM, standard error of the mean.
Figure 1 Bland–Altman 95% limits of agreement plots for the agreement between the various software programmes (provided on the x-axes and y-axes) for CSMA (cm²). The dotted lines are the mean of the difference and the 95% limits of agreement (±2 SD) between the CSMA of reading 1 of observer A and the solid lines of reading 1 of observer B. (A) There was no proportional systematic bias for observer A ($P = 0.908$), whereas there was significant bias for observer B ($P = 0.049$). (B) There was no proportional systematic bias for any observer ($P = 0.738$ and $P = 0.359$). (C) There was no proportional systematic bias for any observer ($P = 0.238$ and $P = 0.704$). (D) There was no proportional systematic bias for any observer ($P = 0.857$ and $P = 0.363$). (E) There was no proportional systematic bias for any observer ($P = 0.185$ and $P = 0.228$). (F) There was no proportional systematic bias for any observer ($P = 0.289$ and $P = 0.843$).
similarity coefficients for the inter-observer and intra-observer comparisons are summarized in Table 5 and depicted in Supporting Information, Figures S3 and S4. All remaining worst Jaccard similarity coefficients are provided in Supporting Information, Figure S5.

The classification of sarcopenia and visceral obesity

The inter-software Cohen’s κ’s of the first reading of observer A for the classification of sarcopenia were 0.96 (between FatSeg and Osirix, Osirix and ImageJ, and ImageJ and sliceOmatic), 0.92 (between FatSeg and ImageJ, and Osirix and sliceOmatic), and 0.88 (between FatSeg and sliceOmatic). No inter-software differences were found in the classification of patients with and without sarcopenia for observer B.

According to the cut-off values used, all software programmes diagnosed sarcopenia in 16 men (55.2%) and 8 women (38.1%). This resulted in a Cohen’s κ of 1.00 for all comparisons between software programmes (P < 0.001).

The Cohen’s κ for the intra-observer agreement of sarcopenia assessment of observer A was 0.96 using FatSeg and ImageJ and 1.00 for Osirix and sliceOmatic (all P < 0.001).

The Cohen’s κ for the inter-observer agreement (reading 1 of observer A vs. reading 1 of observer B) of sarcopenia assessment was 0.92 for sliceOmatic, 0.96 for FatSeg and ImageJ, and 1.00 for Osirix (all P < 0.001).

The classification of visceral obesity

In total, 17 men (58.6%) and 9 women (42.9%) were classified as visceral obese using FatSeg, Osirix, ImageJ, and sliceOmatic in all readings. This resulted in a Cohen’s κ of 1.00 for all comparisons (all P < 0.001).

Discussion

This study shows that the inter-software agreement was excellent for all software programmes. Furthermore, the inter-observer and intra-observer agreements were excellent for four distinct software programmes to assess CSMA, VAT, and SAT on abdominal CT scans with high Jaccard similarity coefficients.

Body composition analyses using abdominal CT scans are increasingly being performed. In multiple surgical populations, such as vascular, gastrointestinal, urological, gynaecological, and transplantation surgery, the association between low skeletal muscle mass and an increased risk of post operative complications, recurrent disease, or impaired survival has been shown. Low skeletal muscle mass is also related to discharge destination in elderly
Figure 2  Jaccard similarity coefficients (lowest and highest are shown) for inter-software comparisons of CSMA, VAT, and SAT (cm$^2$) measurements (reading 1 of observer B). The green area represents similarity, whereas the red area represents discrepancy in measurements. (A) The CSMA measured with FatSeg and ImageJ (1) and FatSeg and sliceOmatic (2), resulting in Jaccard similarity coefficients of 0.998 and 0.927, respectively. (B) The VAT measured with ImageJ and sliceOmatic (1, 2), resulting in Jaccard similarity coefficients of 0.998 and 0.855, respectively. (C) The SAT measured with ImageJ and sliceOmatic (1) and OsiriX and sliceOmatic (2), resulting in Jaccard similarity coefficients of 0.999 and 0.923, respectively.

trauma patients, associated with an increased risk of dose-limiting chemotherapy toxicity and with morbidity and mortality in various oncologic populations, such as lung cancer and melanoma patients. Furthermore, CT-assessed visceral obesity is associated with worse short-term and long-term outcome in distinct patient populations undergoing surgery. Various software programmes have been used to measure body composition in these studies. The current study shows that the results of these studies can reliably be compared. Based on our findings, it is likely that this is also true for other software programmes, which similarly compute skeletal muscle area by quantifying selected voxels within preset HU intensity thresholds (e.g. studies that used software programmes designed in MATLAB (MathWorks, Natick, MA, USA)).

Software programmes for various body composition measurements on CT images, such as adipose tissue surface area,
skeletal muscle tissue surface area, and liver volumetric measurements, have been compared in multiple previous studies, demonstrating high levels of agreement.50–53 Excellent agreement levels between sliceOmatic and ImageJ,54 as well as between observers using sliceOmatic55 for CSMA measurements have previously been reported. Furthermore, excellent agreement levels between OsiriX and ImageJ have been observed for paraspinal muscle measurements on magnetic resonance images.56 Nevertheless, this is the first study to compare multiple software programmes for the measurement of CSMA, VAT, and SAT, showing that previous studies investigating the association between skeletal muscle mass on the one side and visceral or subcutaneous adipose tissue on the other side, and patient outcomes can reliably be compared.

The skeletal muscle area (cm²) measured at a single cross-sectional CT image at the level of the third lumbar vertebra (L3) is linearly related to total body skeletal muscle mass57 and is therefore corrected for height squared (m²), as is conventional for body composition measures. This results in the L3 muscle index (cm²/m²).18 Another frequently used method is measuring the total psoas area (TPA).27 The principle of TPA measurements is identical to L3 muscle area measurements, using single cross-sectional CT images. Therefore, the findings of this study may be extrapolated to TPA measurements as well. Nevertheless, this should be confirmed in a future study.

Significant differences were observed between the mean skeletal muscle areas within and between observers. However, these mean differences are small and consequently could be considered as not clinically relevant. Differences in individual measurements resulted, for instance, from the incorrect annotation of skeletal muscle tissue (see Figure 2A2 for an example of an intra-observer difference). However, we decided not to correct measurements in retrospect to show inter-observer and intra-observer agreements. In our opinion, this study reflects daily practice, with observers who have excellent (radiological) anatomical knowledge performing body composition measurements. Regardless of these human errors and some inter-observer and intra-observer differences, high comparability between software programmes was observed.

Significant differences between VAT measurements were also observed with greater mean differences between software. This could be due to the greater complexity of the measurement technique, as intraluminal content (i.e. fat in stool) needs to be manually erased. The greatest significant mean differences in SAT could partly been explained by the fact that in OsiriX the cutaneous adipose tissue in not included in the SAT, in contrast to the other software programmes. Furthermore, every tissue of interest needs to be manually selected in OsiriX, in contrast to the other programmes in which methods of delineating or a painting brush can be used to select regions of interest.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Mean Jaccard indices for inter-observer (reading 1 of observer A vs. reading 1 of observer B) and intra-observer comparisons (reading 1 vs. reading 2 of observer A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CSMA, cross-sectional muscle area; VAT, visceral adipose tissue area; SAT, subcutaneous adipose tissue area.</td>
</tr>
<tr>
<td></td>
<td><strong>CSMA</strong></td>
</tr>
<tr>
<td></td>
<td>FatSeg</td>
</tr>
<tr>
<td></td>
<td>OsiriX</td>
</tr>
<tr>
<td></td>
<td>ImageJ</td>
</tr>
<tr>
<td></td>
<td>sliceOmatic</td>
</tr>
</tbody>
</table>

DOI: 10.1002/jcsm.12158
Conclusions

In conclusion, this study showed that four different software programmes have an excellent agreement to measure VAT and SAT, and CSMA in particular on abdominal CT scans, which enables reliable comparison of results of studies that use these different software programmes. Multiple slice analysis is preferred for VAT and SAT measurements.

Acknowledgements

The authors would like to thank Laurens Groenendijk and Elsaline Rijkse of the Imaging Trial Office, department of Radiology and Nuclear Medicine, Erasmus MC University Medical Centre, Rotterdam, the Netherlands, for anonymizing and providing the CT scans, and Yves Martel of sliceOmatic (TomoVision, Magog, Canada) for providing a temporary free license to use the software package. The authors certify that they comply with the ethical guidelines for publishing in the Journal of Cachexia, Sarcopenia and Muscle: update 2015.58

JLAvV designed the study, performed the measurements (observer B), analyzed and interpreted the data, and wrote the manuscript. SL designed the study, interpreted the data, and wrote the manuscript. AG performed the measurements (observer A) and analyzed the data. MK and WJN developed and provided the FatSeg software programme and provided technical advice and support for the measurements and the calculation of the Jaccard similarity coefficients. SPW provided statistical and methodological advice and interpreted the data. JWA Burger provided the clinical data. RWFdB and JNMJ interpreted the data and supervised the study. All authors critically revised the manuscript and approved the manuscript for publication.

Online supplementary material

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Supplementary figure 1. Bland-Altman 95% limits of agreement plots for the agreement between the various software programs (provided on the X-axes and Y-axes) for VAT (cm²). The dotted lines are the mean of the difference and the 95% limits of agreement (±2 SD) between the VAT of reading 1 of observer A and the solid lines of reading 1 of observer B

Supplementary figure 1a. There was proportional systematic bias for both observers (p = 0.004 and p = 0.043, respectively).

Supplementary figure 1b. There was proportional systematic bias for observer A (p = 0.038), but not for observer B (p = 0.154).

Supplementary figure 1c. There was proportional systematic bias for both observers (both p = 0.002).

Supplementary figure 1d. There was proportional systematic bias for observer A (p = 0.045), whereas there was no proportional systematic bias for observer B (p = 0.202).
Supplementary figure 1e. There was no proportional systematic bias for any observer (p = 0.412 and p = 0.114, respectively).

Supplementary figure 1f. There was no proportional systematic bias for observer A (p = 0.068), whereas there was significant bias for observer B (p = 0.014).

Supplementary figure 2. Bland-Altman 95% limits of agreement plots for the agreement between the various software programs (provided on the X-axes and Y-axes) for SAT (cm²). The dotted lines are the mean of the difference and the 95% limits of agreement (±2 SD) between the SAT of reading 1 of observer A and the solid lines of reading 1 of observer B.

Supplementary figure 2a. There was no proportional systematic bias for any observer (p = 0.534 and p = 0.801, respectively).

Supplementary figure 2b. There was no proportional systematic bias for any observer (p = 0.538 and p = 0.112, respectively).

Supplementary figure 2c. There was proportional systematic bias for observer A (p = 0.031), whereas there was no proportional systematic bias for observer B (p = 0.134).

Supplementary figure 2d. There was no proportional systematic bias for any observer (p = 0.853 and p = 0.344, respectively).

Supplementary figure 2e. There was no proportional systematic bias for any observer (p = 0.511 and p = 0.305, respectively).

Supplementary figure 2f. There was no proportional systematic bias for any observer (p = 0.175 and p = 0.939, respectively).

Supplementary figure 3. Jaccard similarity coefficients (lowest and highest are shown) for inter-observer comparisons of CSMA, VAT, and SAT (cm²) measurements (reading 1 of observer A versus reading 1 of observer B). The green area represents similarity, whereas the red area represents discrepancy in measurements.

Supplementary figure 3a. The CSMA measured with FatSeg (1) and ImageJ (2), resulting in Jaccard similarity coefficients of 0.997 and 0.931, respectively.

Supplementary figure 3b. The VAT measured with FatSeg (1) and OsiriX (2), resulting in Jaccard similarity coefficients of 0.998 and 0.835, respectively.

Supplementary figure 3c. The SAT measured with OsiriX (1) and ImageJ (2), resulting in Jaccard similarity coefficients of 1.000 and 0.899, respectively.

Supplementary figure 4. Jaccard similarity coefficients (lowest and highest are shown) for intra-observer comparisons of CSMA, VAT, and SAT (cm²) measurements (reading 1 versus reading 2 of observer A). The green area represents similarity, whereas the red area represents discrepancy in measurements.

Supplementary figure 4a. The CSMA measured with FatSeg (1) and OsiriX (2), resulting in Jaccard similarity coefficients of 1.000 and 0.953, respectively.

Supplementary figure 4b. The VAT measured with FatSeg (1) and OsiriX (2), resulting in Jaccard similarity coefficients of 0.999 and 0.838, respectively.

Supplementary figure 4c. The SAT measured with ImageJ (1, 2), resulting in Jaccard similarity coefficients of 1.000 and 0.900, respectively. Furthermore, a Jaccard similarity coefficient of 1.000 was also observed in two other patients measured with FatSeg and OsiriX.

Supplementary figure 5. All worst Jaccard similarity coefficients for the inter-software (5a), inter-observer (5b) and intra-observer (5c) agreement for the three body composition analyses except the ones that are shown in figure 2 and supplementary figures 2 and 3.

Supplementary figure 5a1a. CSMA measured with FatSeg and ImageJ resulting in a Jaccard similarity coefficient of 0.959.

Supplementary figure 5a1b. CSMA measured with FatSeg and OsiriX resulting in a Jaccard similarity coefficient of 0.948.

Supplementary figure 5a1c. CSMA measured with ImageJ and OsiriX resulting in a Jaccard similarity coefficient of 0.948.

Supplementary figure 5a1d. CSMA measured with ImageJ and sliceOmatic resulting in a Jaccard similarity coefficient of 0.932.

Supplementary figure 5a1e. CSMA measured with OsiriX and sliceOmatic resulting in a Jaccard similarity coefficient of 0.944.

Supplementary figure 5a2a. VAT measured with FatSeg and ImageJ resulting in a Jaccard similarity coefficient of 0.903.

Supplementary figure 5a2b. VAT measured with FatSeg and OsiriX resulting in a Jaccard similarity coefficient of 0.886.

Supplementary figure 5a2c. VAT measured with FatSeg and sliceOmatic resulting in a Jaccard similarity coefficient of 0.860.

Supplementary figure 5a2d. VAT measured with ImageJ and OsiriX resulting in a Jaccard similarity coefficient of 0.891.

Supplementary figure 5a2e. VAT measured with OsiriX and sliceOmatic resulting in a Jaccard similarity coefficient of 0.884.

Supplementary figure 5a3a. SAT measured with FatSeg and ImageJ resulting in a Jaccard similarity coefficient of 0.968.

Supplementary figure 5a3b. SAT measured with FatSeg and OsiriX resulting in a Jaccard similarity coefficient of 0.926.

Supplementary figure 5a3c. SAT measured with FatSeg and sliceOmatic resulting in a Jaccard similarity coefficient of 0.960.

Supplementary figure 5a3d. SAT measured with ImageJ and OsiriX resulting in a Jaccard similarity coefficient of 0.927.

Supplementary figure 5a3e. SAT measured with ImageJ and sliceOmatic resulting in a Jaccard similarity coefficient of 0.965.

Supplementary figure 5b1a. CSMA measured with FatSeg resulting in a Jaccard similarity coefficient of 0.949.

Supplementary figure 5b1b. CSMA measured with OsiriX resulting in a Jaccard similarity coefficient of 0.931.

Supplementary figure 5b1c. CSMA measured with sliceOmatic resulting in a Jaccard similarity coefficient of 0.939.

Supplementary figure 5b2a. VAT measured with FatSeg resulting in a Jaccard similarity coefficient of 0.908.

Supplementary figure 5b2b. VAT measured with ImageJ resulting in a Jaccard similarity coefficient of 0.905.

Supplementary figure 5b2c. VAT measured with sliceOmatic resulting in a Jaccard similarity coefficient of 0.876.

Supplementary figure 5b3a. SAT measured with FatSeg resulting in a Jaccard similarity coefficient of 0.969.

Supplementary figure 5b3b. SAT measured with OsiriX resulting in a Jaccard similarity coefficient of 0.961.

Supplementary figure 5b3c. SAT measured with sliceOmatic resulting in a Jaccard similarity coefficient of 0.959.

Supplementary figure 5c1a. CSMA measured with FatSeg...
resulting in a Jaccard similarity coefficient of 0.961.

**Supplementary figure 5c1b.** CSMA measured with ImageJ resulting in a Jaccard similarity coefficient of 0.948.

**Supplementary figure 5c1c.** CSMA measured with sliceOmatic resulting in a Jaccard similarity coefficient of 0.961.

**Supplementary figure 5c2a.** VAT measured with FatSeg resulting in a Jaccard similarity coefficient of 0.916.

**Supplementary figure 5c2b.** VAT measured with ImageJ resulting in a Jaccard similarity coefficient of 0.891.

**Supplementary figure 5c2c.** VAT measured with sliceOmatic resulting in a Jaccard similarity coefficient of 0.901.

**Supplementary figure 5c3a.** SAT measured with FatSeg resulting in a Jaccard similarity coefficient of 0.956.

**Conflict of interest**

J.L.A. van Vugt, S. Levoiger, A. Gharbarhan, M. Koek, J.W.A. Burger, S.P. Willemsen, R.W.F. de Bruin and J.N.M. U Izermanns have no conflicts of interest. W.J. Niessen is co-founder and shareholder of Quantib BV.

**References**


Software programmes for body composition measurements on CT

297


