ORIGINAL RESEARCH

Computed tomography volume measurements of the liver using a liver segmentation and analysis package: an intra- and inter-rater reliability study

Jessie T Childs∗1, Kerry A Thoirs2, Adrian J Esterman3

1University of South Australia, Adelaide, Australia
2International Centre of Allied Health Research, University of South Australia, Adelaide, Australia
3FACE, University of South Australia, Adelaide, Australia and Centre for Research Excellence in Chronic Disease Prevention, James Cook University, Cairns, Australia

Received: May 27, 2015 Accepted: July 2, 2015 Online Published: July 28, 2015
DOI: 10.5430/jbgc.v5n2p17 URL: http://dx.doi.org/10.5430/jbgc.v5n2p17

Abstract

Introduction: Recent advances in technology have seen the introduction of software assisted image post-processing (SAIP) tools to calculate the volume of the liver from Computed Tomography (CT) images. One such SAIP tool is the semi-automated Philips liver segmentation and analysis package (Endhoven, The Netherlands). The intra- and inter-rater reliability of the liver volumes of 16 participants calculated using this tool was assessed.

Methods: Two CT Technologists accessed Abdominal CT data sets and calculated the volume of the liver using the Philips liver segmentation and analysis package. One technologist repeated their calculation.

Results: Results demonstrated high intra-rater reliability (ICC 0.9997) and high inter-rater reliability (ICC 0.99994). Analysis with Bland Altman plots showed the mean bias for both intra- and inter-rater reliability to be close to zero with acceptable limits of agreement.

Conclusions: This study has shown that liver volume measurements using the Philips SAIP (Endhoven, The Netherlands) can be taken reliably by CT technologists, with confidence that the same results would be obtained between different CT technologists.

Key Words: Computed tomography, Liver, Reliability, Software assisted image post-processing, Volume

1 Introduction

The ability to measure the size of the adult liver improves patient outcomes by aiding in surgical planning and follow-up, disease diagnosis and tracking of disease progression or response to treatment over time.[1-4] For patients undergoing major hepatic resection, the likelihood of post – hepatic liver failure is three times higher in patients who are left with < 25% of their initial liver volume compared to those who are left with ≥ 25% of their initial liver volume.[5] Mortality and morbidity rates for major hepatic resection have been reported to be as high as 30% with post hepatic liver failure being the major cause. Therefore, knowledge of the pre and post-surgical liver volumes are essential for superior patient outcomes.[6] In some instances, hepatic resection is performed in two separate surgeries. Appropriate increase in liver volume 6 days following initial surgery enables the second part of the surgery to be safely performed.[7] Liver volume is also commonly assessed 3-
4 weeks following portal vein embolization, which is done as part of resection, to assess for appropriate liver hypertrophy. An increase in size of 5% is associated with improved patient outcomes. In addition to this, liver volume calculations are commonly used to estimate drainage volume prior to endoscopic biliary drainage. Liver volume is also commonly assessed six weeks and one week prior to gastric banding surgery to ensure adequate volume reduction.

Computed Tomography (CT) is currently considered the gold standard for assessment of the volume of the adult liver. Magnetic Resonance Imaging (MRI) has also been shown to be accurate in taking volume measurements of the liver. Despite the advantage MRI offers in the form of not using ionising radiation, in comparison CT is more readily available, faster and less expensive to perform than MRI.

Prior to the introduction of automated methods of CT liver measurement, measurement of liver volume from CT images was performed using manual volumetry, a technique performed by manually tracing the liver boundary on each CT slice to determine the volume of liver on each slice. The individual slice volumes were then summed to calculate an overall liver volume. Recent advances in CT imaging technology have seen the introduction of software programs that automatically calculate liver volume; a less operator dependent method. Multiple automated and semi-automated liver segmentation tools have been developed and are collectively referred to as software assisted image post-processing (SAIP) tools.

SAIP tools are increasing in popularity and have been proven to be very accurate in predicting liver volumes. Studies have shown that measurement of liver volume using SAIP tools correlate well with manual volumetry and with the added advantage of being a substantially faster method of volume calculation, they are being touted as the new gold standard for liver volumetry.

The purpose of this study was to investigate the intra- and inter-rater reliability of a specific SAIP tool, namely, the Philips liver segmentation and analysis package (Eindhoven, The Netherlands). The reliability testing of this tool was undertaken as a preliminary investigation to establish the reliability of two observers who were to take part in a larger study. This Philips liver segmentation and analysis package (Eindhoven, The Netherlands) provides an automatic volume calculation of the liver by highlighting the liver tissue within each slice (instead of requiring the technologists to hand trace the slices as seen in manual volumetry) and then summing the individual slice volumes. A CT operator then provides a subjective component to the calculation by accessing the individual slices and manually adding or removing highlighted tissue which may have been missed by the software or included in error.

CT operator intra- and inter-rater reliability for volume measurements of the liver using the Philips liver segmentation and analysis package (Eindhoven, The Netherlands) has not been published. Reliability testing of this SAIP tool will inform researchers and clinicians of the magnitude of variance of liver volume measurements that can be attributed to any imprecision of this measurement technique.

## 2 Methods

Ethics approval was sought and granted from the ethics committee of the University of South Australia.

A sample size calculation was performed using PASS 11 (NCSS, Utah USA) which determined a sample size of 16 participants was required with 2 raters per participant to achieve 80% power to detect an alternative hypothesis, intra-class correlation (ICC) of 0.7 against the null hypothesis that the ICC is 0.2, using an independent samples t-test with a significance level of 0.05.

The 16 participants were recruited as a sample of convenience from the CT patient population that had been referred by their General Medical Practitioner to a private radiology firm for an abdominal CT as part of their medical care. An information sheet was provided to potential participants and once they had agreed to take part, written consent was obtained prior to their CT. Participants were excluded if the CT they had been referred for would not image their entire liver, if they were unable to read and comprehend the information sheet, or if they were under 18 years of age.

There were nine female and seven male participants, mean age (standard deviation) 62 (16) years (range 24 – 89 years). Participants were not asked questions regarding their medical history. A formal radiology report for each scan was prepared by the onsite radiologist and returned to the participant’s general medical practitioner.

The examinations were undertaken by one of two qualified CT technologists (SR & BW) on a Philips Ingenuity 128 slice low dose CT machine with iDose 4 software (Philips Medical Systems, Best, Netherlands). A standard adult abdominal imaging protocol was used with the patient lying supine on a CT table in a state of respiratory inspiration during image collection. The studies were a combination of contrast enhanced and non-contrast depending on the clinical question being asked. kV was set at 120, mA was set at a preset Philip DoseRight Index of 16, collimation 64 × 0.625, pitch 1.172, rotation time 0.4 seconds with a gantry angle of 0 degrees. The slice thickness used was 1.0 mm with a slice increment of 0.5 mm. A standard B filter was used with iDose level 4, render mode set to average and a 512 matrix. The participant’s images were saved to the CT workstation hard drive.

At a date no sooner than two weeks following the scan, both technologists accessed the CT images separately and performed the measurements from the thin slice data using the liver segmentation and analysis software (Eindhoven, The Netherlands).
Figure 1 depicts a CT image where the software has identified and highlighted the liver tissue; guided by liver segments and surrounding vascular landmarks.

The results of the intra- and inter-rater reliability testing are shown in Table 1 and the results of the statistical analyses are shown in Table 2. The time taken for each volume calculation was recorded as being less than 5 minutes in all cases.

Figure 2: (A) A CT slice with some liver that has been omitted from the cross sectional area measurement and a part of the heart that has been incorrectly included. (B) The previously omitted liver tissue has now been included in the measurement (marked in blue). (C) Using the erase tool, the incorrectly included heart tissue is erased. (D) The liver is now all correctly included and the heart tissue deleted from the measurement.

Intraclass correlation coefficients are usually interpreted as follows: values of 0.7 and below are considered as having poor agreement. Values of 0.7-0.8 are considered to have strong agreement, and values of 0.81 and above are considered as having very strong agreement. The results of 0.997 and 0.99994 show near perfect agreement for both intra- and inter-rater reliability respectively.
Table 1: Total liver volumes (cc)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Observer 1 (SR)</th>
<th>Observer 1 (SR) Repeat Measurement</th>
<th>Observer 2 (BW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1515.6</td>
<td>1514.2</td>
<td>1531.1</td>
</tr>
<tr>
<td>2</td>
<td>1878.9</td>
<td>1884.6</td>
<td>1893.9</td>
</tr>
<tr>
<td>3</td>
<td>1413.2</td>
<td>1419.5</td>
<td>1416.9</td>
</tr>
<tr>
<td>4</td>
<td>1730.7</td>
<td>1730.2</td>
<td>1739.6</td>
</tr>
<tr>
<td>5</td>
<td>1265.2</td>
<td>1263.4</td>
<td>1249.1</td>
</tr>
<tr>
<td>6</td>
<td>1691.1</td>
<td>1717.7</td>
<td>1710.9</td>
</tr>
<tr>
<td>7</td>
<td>1453.3</td>
<td>1458.6</td>
<td>1467</td>
</tr>
<tr>
<td>8</td>
<td>1003</td>
<td>1000.5</td>
<td>1006.5</td>
</tr>
<tr>
<td>9</td>
<td>1859.4</td>
<td>1844.7</td>
<td>1842.5</td>
</tr>
<tr>
<td>10</td>
<td>1470.4</td>
<td>1471.5</td>
<td>1468.3</td>
</tr>
<tr>
<td>11</td>
<td>2670.4</td>
<td>2660.8</td>
<td>2653.2</td>
</tr>
<tr>
<td>12</td>
<td>2154.6</td>
<td>2147.3</td>
<td>2168.2</td>
</tr>
<tr>
<td>13</td>
<td>1342.4</td>
<td>1335.3</td>
<td>1331.5</td>
</tr>
<tr>
<td>14</td>
<td>1620.4</td>
<td>1610.9</td>
<td>1597</td>
</tr>
<tr>
<td>15</td>
<td>1466.7</td>
<td>1464.3</td>
<td>1461.8</td>
</tr>
<tr>
<td>16</td>
<td>1387.3</td>
<td>1377</td>
<td>1379.5</td>
</tr>
</tbody>
</table>

Table 2: Results of statistical analysis

<table>
<thead>
<tr>
<th>Statistical Tool</th>
<th>Intra-rater reliability of liver volume measurements</th>
<th>Inter-rater reliability of liver volume measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC (95% confidence intervals)</td>
<td>0.99997 (0.99992 to 0.99999)</td>
<td>0.999994 (0.9983 to 0.9998)</td>
</tr>
<tr>
<td>Mean Bias</td>
<td>1.4cc</td>
<td>0.4cc</td>
</tr>
<tr>
<td>Limits of Agreement</td>
<td>37.9cc</td>
<td>54.7cc</td>
</tr>
<tr>
<td>Pattern of Bias</td>
<td>Slight trend of error increasing with increasing liver size</td>
<td>Increasing error with increasing liver size</td>
</tr>
<tr>
<td>Significance t-test</td>
<td>0.571</td>
<td>0.100</td>
</tr>
</tbody>
</table>

4 Discussion

In order to be useful in disease tracking, surgical planning and treatment monitoring liver volumes need not only to be accurate, but also to be reproducible and reliable. This is particularly important when the method of calculation involves subjective operator involvement. While SAIP tools have been demonstrated to have high accuracy, no literature could be found on the reliability of CT liver volumes calculated using semi-automated SAIP's. However, Sandrasegaran et al. 1999 [25] demonstrated high reliability for manual volumetry of the liver (ICC = 0.999 intra-rater and ICC = 0.997 inter-rater) on 35 subjects. Whilst no comment can be made for other SAIP tools, this particular tool is simple to use and requires limited training. A short session of approximately 5 minutes is all that is required. One of the raters in this study had been using the tool for approximately 6 months prior to the beginning of the study, and one was given a five minute training session just prior to the beginning of the study. The limited training required, and the fact that results were reliable between a technologist experienced in using the package and a technologist who was a novice at using the package, indicates that reliability is achievable independent of operator experience.

Multiple studies have compared SAIP tools to manual volumetry. Suzuki et al. 2011 [15] compared two different SAIP tools to the manual volumetry method. One was a proprietary Volume Tracing Advanced Vessel Analysis Tool (Philips Healthcare) and the other, an SAIP tool developed by the authors which was based on Geodesic active contour segmentation. This study found that not only did the two SAIP tools correlate well with manual volumetry but that they were significantly faster: providing volume estimates at less than one minute per case compared to manual volumetry which took on average 35-40 minutes per case. Similar results in regards to both the accuracy of measurements and faster measurement times are echoed in the studies by Shin et al. 2013 [26] and Nakayama et al. 2006 [27]. Our study also showed semi-automated volume calculations to be faster than reported manual volumetry calculations with all calculations being recorded as taking less than 5 minutes. It would follow that if SAIP tools have been shown to be accurate compared to manual volumetry and manual volumetry has been shown to be reliable, that SAIP tools would also be reliable. This study has indeed confirmed this to be the case, by demonstrating high intra- and inter-rater reliability. The CT technologists in this study have over 10 years of experience, but they were both novices at using the semi-
automated Philips liver segmentation and analysis package (Endhoven, The Netherlands). This asserts that consistent volume measurements of the liver can be achieved using this software even where operators have limited experience in the technique.

Liver volume assessed 3-4 weeks post portal vein embolisation shows an increase in volume of 5% being associated with improved patient outcomes. Lewis et al. 2006[10] found liver volume reduced by 14% over a six week period following the implementation of a low calorie diet. Edholm et al. 2014[28] showed an average decrease in liver size of 18% in ten female participants following two weeks on a low calorie diet. Van Theil et al. 1987[29] evaluated the size of two recipient transplant liver patients and showed the liver size in both recipients to increase by an average of 70 cc per day. Jeschke et al. 2007[30] studied 102 children exposed to thermal injury and found their liver size increased on average by 85% in the first two weeks following injury.

With respect to limitations, this study only examined the reliability of the Philips SAIP software (Eindhoven, The Netherlands), although it is likely that similar results would be found for other SAIP tools. The use of only two raters is arguably a potential source of bias, nonetheless it is common practice, and many other papers use only two raters to confirm reliability.[31, 32] A further limitation of this study is the sample did not contain any participants with resected livers, which is a population who would be examined clinically using this technique.

5 Conclusion

The advantage of SAIP tools lies in their ability to quickly and accurately calculate liver volume automatically with only fine adjustment required by the operator. This study has shown that liver volume measurements of both normal and abnormal livers using the Philips SAIP (Endhoven, The Netherlands) can be taken reliably by CT technologists, with confidence that the same results would be obtained between different CT technologists.

Acknowledgements

The authors would like to acknowledge the staff and management of Sound Radiology, Adelaide, South Australia for the use of their equipment and assistance with this study. The authors would also like to acknowledge Philips Healthcare for the provision of the Liver Segmentation Software Package.

References


y.org/10.1002/lt.23461

http://dx.doi.org/10.1002/10.1016/j.1600-0676.1984.tb00940.x


