A Comparison of Radiographers and Radiologists in CT based measurements of Abdominal Aortic Aneurysms.

Mr Andrew England¹, B.Sc (Hons), M.Sc

Miss Abigail Best¹

Miss Charlotte Friend¹

Directorate of Medical Imaging & Radiotherapy¹

University of Liverpool, Johnston Building

Quadrangle, Brownlow Hill, Liverpool, L69 3GB, United Kingdom

Tel:+441517945123    Fax:+441517945719

Corresponding author:  Mr Andrew England, Directorate of Medical Imaging & Radiotherapy, University of Liverpool, (a.england@liv.ac.uk).
ABSTRACT

AIM: To evaluate the variability of CT AAA measurements undertaken by radiologists and radiographers.

METHODS: 19 observers (4 radiologists, 15 radiographers) were invited to independently measurement maximum aneurysm diameter (Dmax) on ten CT scans. Each CT scan was presented randomly to each observer; four were duplicates in order to test intra-observer variability. All measurements were taken from axial CT images using electronic callipers, all observers were blinded to any previous measurements. Both the slice number and the maximum AAA diameter in any plane were recorded.

RESULTS: Intra-observer variability was lower for radiographers with a mean paired difference of -0.18 ± 2.6 mm compared to -2.1 ± 3.5 mm (P=0.054). Inter-observer variability within each observer group was comparable, radiographers 0.1 ± 5.0 mm; radiologists -0.1 ± 3.1 mm (P=0.680). When directly comparing between the two groups mean difference was -2.0 ± 4.0 mm with 43% of paired measurements ≤2mm or less and 78% ≤5mm. Slice selection was less variable between the two groups with 88% of repeat radiographer measurements within +/- 1 slice and 91% of radiologists measurements with +/- 1 slice (P=0.228).

CONCLUSION: The accuracy of radiographers in performing AAA CT measurements is comparable to radiologists. Variability exists for both professions and in some instances may be clinical significant, observers should be aware of measurement variability issues and have an understanding of the factors responsible. Careful and repeat measurements of AAAs around 5.5 cm are recommended in order to define treatment.
Introduction

Both ultrasound and CT have been used for many years to determine the size of abdominal aortic aneurysms (AAAs). The major problem with AAAs is the risk of rupture; AAA’s are often asymptomatic and if left undetected they will continue to expand and eventually rupture. Decisions about the treatment of AAAs are traditionally based upon the maximum cross-sectional diameter. If the aneurysm is 5.5cm or larger in diameter intervention is generally deemed appropriate(1). Aneurysms smaller than this or on the borderline will be deemed of lower rupture risk and may be monitored by ultrasound surveillance(2).

Accurate measurement of the AAA is paramount since if an aneurysm was measured to be significantly smaller than its actual size, the patient would be treated conservatively and exposed to a high risk of rupture. Conversely if the aneurysm is measured to be larger than the actual size then the patient may undergo an unnecessary surgical procedure which carries risks. Furthermore, many clinicians’ are using AAA growth or shrinkage to assess treatment outcomes and thus precise measurements are essential to ensure optimum patient management. Variability of AAA measurements has been reported both between imaging modalities(3) and between observers of different disciplines(4). With radiographic role extension branching into elements of CT image interpretation the aim of this project was to investigate the different between radiographers and radiologists in undertaking CT measurements of AAA’s. Previous research in this area has explored the variability between a single radiologist and radiographer(4) but not for multiple observers.
Methods

A set of five CT scans were randomly selected from a hospital Picture Archiving and Communication System (PACS). All scans were acquired from the same CT scanner (Siemens Somatom Sensation 16, Siemens Medical Systems, Erlangen, Germany) using the same acquisition parameters. All CT datasets were anonymised and transferred onto a laptop computer in DICOM format. Monitor calibration and performance tests were undertaken on the computer to ensure optimum image quality and that this was maintained throughout the study. Each of the CT datasets consisted of approximately sixty slices of the abdominal and demonstrated an AAA with different maximum aneurysm diameters. CT acquisition comprised of 5mm full-field axial CT images of the abdominal aorta following injection of iodinated contrast media with images reconstructed every 5mm. For the purposes of this study images were assessed using the computer software Dicom Works version 1.3.5 (DicomWorks, Lyon, France).

Observers (consultant radiologists or diagnostic radiographers) were recruited from two University Hospitals in the north-west of England. Each observer was given a basic eye test to confirm that eyesight was not a variable. Observer was then asked to scroll down through each CT scan using the scroll wheel on the mouse and carefully choose the slice which depicted the aorta at its maximal diameter. Each observer was then asked to measure and record the maximal diameter of the aorta in any plane (adventitia to adventitia). All measurements were taken using electronic callipers at standard abdominal window levels (400W 40L). The slice number of each measurement was also recorded. With the exception of scan A (which was presented to all observers first), all other scans were presented in a random order. To determine intra-observer variability each of the
scans was duplicated and given an alternative identifier, this increased the number of scans to ten. The first CT measurement (Scan A) from each observer was disregarded in order to allow the observers to become familiar with the study methodology and the DicomWorks computer software. Each observer was provided with the same predefined instructions and was not aware that any of the scans were repeated within the set. Before the study commenced duplicate scans had several initial CT slices removed from the dataset so that the start position varied in order to discourage observers from recognising any duplicates. Duplicate CT scans were presented at a point in the test other than directly before or after its twin to prevent further detection. For all measurements room lighting was kept standard (ambient) and measurements were undertaken in a room which resembled a standard radiological reporting room. Before the study commenced the local research ethics committee approved the study.

**Statistical Analysis**

To assess the intra-observer variability the difference in paired measurements were calculated for repeat measurements of the same CT scan by the same observer. These differences were summarised in mean and standard deviation as well as graphical illustration of the difference made according to a method described by Bland and Altman(6). Inter-observer differences were determined by calculating the difference between each observer and all other observers both within and between observer groups. Again results were displayed in a similar manner to intra-observer variability. All results were recorded on the statistical computer package SPSS 16.0 (SPSS Inc, Chicago, IL). Continuous data was expressed as mean values plus or minus its standard deviation. Inferential statistical
analysis were undertaken in order to identify statistically significant differences. P values of 0.05 or less were considered to be statistically significant.

**Results**

Independent measurements were taken by 19 observers (4 radiologists and 15 radiographers). Analysis of intra-observer variability was undertaken to determine the difference in AAA measurements between individual observers. Intra-observer variability was higher for radiologists -2.1 ±3.5mm compared with -0.18±2.6mm for radiographers. Although differences existed between the two observer types and was maintained when categorising the paired differences (≤ 5mm) this did not reach statistical significance. Full results of the intra-observer analysis for each observer type are shown in Table 1 and Figures 1 and 2.

Inter-observer variability within each observer group was calculated in order to allow comparison of measurements made by observers of the same discipline (Table 2, Figures 3 & 4). Mean paired differences were equivocal for both disciplines (radiologists -0.09±3.1mm; radiographers 0.1±5.0mm) but radiologists had increased consistency when paired differences were categorised. For radiologists 67% of paired measurements were within 2mm and 91% were within 5mm compared to 42% and 80% for radiographers, these results were not statistically significant.
Inter-observer variability analysis directly comparing radiologists' measurements with radiographers are displayed in Table 3 and Figure 5. Inter-observer variability was good with 43% of paired measurements within 2mm and 78% within 5mm. Finally, radiologists and radiographers were compared by the slice selection for each measurement, results demonstrate that repeat slice selection was within 1 CT slice or less for 91% of radiologists and 88% of radiographer measurements (NS, P=0.228; Table 4).

Discussion

The diameter of an abdominal aortic aneurysm (AAA) is a well established objective criterion for selecting both patients for treatment and for assessing the results following endovascular repair. There are currently no reports investigating variability of CT measurements of AAAs between groups of radiologists and radiographers. Our data shows (Table 1) that intra-observer mean differences were lower -0.2 ± 2.6 mm for radiographers when compared to radiologists. This may suggest that the radiographers were internally more reliable and accurately replicated the task when asked to do so. However, when considering the limits of agreement and scattering on Bland-Altman plots (Figs 1 & 2) both observer types appeared to perform equally well. Inferential analysis confirmed that there
was no statistically significant difference in intra-observer variability between radiologists and radiographers. Again when comparing paired measurements within 2mm of less both groups averaged around 60% of measurements within this criterion. Low numbers of observers in the radiologist group (n=4) undoubtedly had a factor in explaining the differences in internal repeat measurements between the groups.

The inter-observer variability within each group using the mean paired difference and SD suggests that both groups performed equally. However, when analysing the number of measurements within 2mm and 5mm radiologists had more paired measurements within each of these categories 67% & 91% compared with 42% and 80%. Further analysis of the Bland Altman plots confirms that the limits of agreement are marginally smaller for the radiologists than the radiographers. Data presented in Table 3 shows that the mean paired difference when directly comparing both groups was 2.0 ± 4.0 mm with four out of five paired measurements being within ≤5mm, this level is generally considered to be the criteria for establishing a AAA size change(7;8). Analysis of Bland Altman plots and limits of agreement against suggests that observer variability when directly assessed between groups is only marginally greater than within groups and between the same observer.

In a study by Lederle et al(7) who compared local CT measurements with a central core laboratory. They found 17% of paired measurements were ≥5 mm which compares well to our study where 22% of measurement pairs were ≥5 mm. What is not clear from this study by Lederle is the training and experience of the core laboratory observers.
In a study by Jaakkola et al.(8) who compared US and CT measurements of AAA, they reported that 62% of paired CT measurements undertaken by radiologists were ≤2 mm of each other. Radiologists in our study compare favourably where 67% of paired measurements were ≤2 mm of each other. There is a single report comparing radiologists with radiographers performed measurements. This report by England et al(4) compared a single radiologist with a single radiographer highlighted a mean difference of 0.56±3.2 mm between the two observers. Data from our study suggests increased variability with mean differences of -2.0 ± 4.0 mm between the two disciplines. These results cannot be directly compared as in England’s original study image analysis was undertaken on a 3D computer workstation and involved only single observer in each category. Evidence within the literature suggests that variability is decreased when using fewer observers(9) and 3D imaging techniques(10,12).

Several comments can be made about this study. Understanding the clinical significance of measurements is an important factor. With 5.5 cm diameters being universally accepted as the treatment size criterion arguments could be raised about the measurements of clearly small or exceptionally large aneurysms. Measurement of disease at these levels could be less meticulous since from a clinical perspective it is not vital that the AAA is measured to the same degree of accuracy. This could be explained by increases in the paired differences for small and large AAAs when reviewing the Bland-Altman plots (Figure 5). These differences were not present in our study and this may reflect the clearly defined instructions to each participant and confirms that our variability was independent of AAA
size. If clear and precise instructions resulted in lower variability then this finding is supported by Cayne et al(10) who concludes that inter-observer variability cannot be eliminated but can be reduced by standardisation of measurement techniques.

Another possible problem with this study could be the effect of the aorta bending within the abdomen. A tortuous aorta could lead to the aorta being imaged obliquely at its maximum point and this would potentially make the AAA much larger and only more experienced observers e.g. radiologists may take this into consideration. For future studies a 3D evaluation of the aorta may show improved variability and this has been further advocated in studies by Broeders et al(12) and by Cayne et al(10). In addition to 3D evaluations Cayne et al(10) further argues for the use of computer-assisted measurements and the assessment of other AAA growth variables e.g. volume when sizing an aneurysm. Arguing against this Abada et al.(11) concludes that maximum diameter measurements are sufficient for sizing AAAs and that volume measurements should only be used in cases where diameters do not allow a clear classification. What is clear from all of these publications is that adequate consideration of the effect of observer variability when performing AAA measurements is paramount especially when deciding treatment and defining measurement protocols.

The measurement of maximum AAA diameter on CT scans generates two tasks 1) identification of the slice where the aorta is depicted at its maximum diameter, in our study this was from a series of around 60 images and 2) selecting two points on an image which
demonstrate the maximum dimensions of the AAA which forms the measurement.

Variability is almost certainly due to both and will be affected by other external factors such as observer experience and contrast sensitivity. From our data (Table 4) it would appear that variability is less influenced by slice selection and more by the decisions around plotting the measurement points from the axial images. From discussions with participants it would also appear that certain CT datasets are harder to measure than others. One reason for this is the ability of an observer to select the boundaries of the aorta. This can be challenging when there is no clear separation of the aortic wall from surrounding structures e.g. small bowel. This factor will vary from patient to patient but could be dependent on an individual’s ability to identify junctions between low contrast structures. A further study is planned which will assess the effect of visual contrast sensitivity on the ability to perform accurate CT based AAA measurements.

Role extension within radiographic practice is now well established, despite this there is still a lack of evidence surrounding the accuracy of radiographers in undertaking measurements from medical images. At present we feel that there is good agreement between radiologists and radiographers in performing aortic CT measurements. Practically to understand whether this has any potential within NHS radiology departments was beyond the scope of this study. Any initiative which proposes a change to practice would need a thorough evaluation of the benefits and risks and should be encouraged if there are intended patient benefits.
Conclusion

A good level of agreement exists between radiologists and radiographers in performing CT measurements of maximum AAA diameter. Variability for both professions does exist and can be significant in certain situations, observers should be aware of the existence of variability especially when making treatment decisions. It is technically feasible for radiographers to perform such measurements, where this area of role extension should be explored needs further investigation. Understanding the factors which play a role in observer variability is paramount; variability may be decreased if using standardised measurement protocols, 3D techniques and computer-assisted measurements. If the later is to be accepted then these measurements will require validation and clinical checking before deciding treatment.
ACKNOWLEDGEMENTS

The authors would like to thank all study participants for their help in completing this study. Furthermore, the authors wish to thank the Society & College of Radiographers for awarding a CoRIPS research grant to support this project.

LEGENDS FOR FIGURES

Figure 1. Bland-Altman plot demonstrating intra-observer variability between radiographers.

Figure 2. Bland-Altman plot demonstrating intra-observer variability between radiologists.

Figure 3. Bland-Altman plot demonstrating inter-observer variability between radiographers.

Figure 4. Bland-Altman plot demonstrating inter-observer variability between radiologists.

Figure 5. Bland-Altman plot demonstrating inter-observer variability between radiographers and radiologists.
Reference List


(5) http://ludens.cl/photo/montest.html.


Figure 4
Table 1. Intra-observer variability when measuring maximum AAA diameter (Dmax).

<table>
<thead>
<tr>
<th>Observer type</th>
<th>Paired Measurements (n)</th>
<th>Mean paired difference (mm)</th>
<th>SD (mm)</th>
<th>Frequency of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologists</td>
<td>16</td>
<td>-2.1</td>
<td>3.5</td>
<td>10 (63)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11 (69)</td>
</tr>
<tr>
<td>Radiographers</td>
<td>60</td>
<td>-0.18</td>
<td>2.6</td>
<td>36 (60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56 (93)</td>
</tr>
</tbody>
</table>

*Numbers in parenthesis highlights the percentage for each category. All continuous data is expressed in millimetres unless otherwise stated. SD, standard deviation; Dmax, maximum AAA diameter. Inferential analysis between groups using the student t-test demonstrated no statistically significant difference (P=0.054).

Table 2. Inter-observer variability within each observer group when measuring maximum AAA diameter (Dmax).

<table>
<thead>
<tr>
<th>Observer type</th>
<th>Observed (n)</th>
<th>Mean paired difference (mm)</th>
<th>SD (mm)</th>
<th>Frequency of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>54</td>
<td>-0.09</td>
<td>3.1</td>
<td>36 (67%)</td>
</tr>
<tr>
<td></td>
<td>945</td>
<td>0.10</td>
<td>5.0</td>
<td>396 (42%)</td>
</tr>
</tbody>
</table>

*Numbers in parenthesis highlights the percentage for each category. All continuous data is expressed in millimetres unless otherwise stated. SD, standard deviation; Dmax, maximum AAA diameter. Inferential analysis between groups using the student t-test demonstrated no statistically significant difference (P=0.680).
Table 3. Inter-observer variability between observer groups when measuring maximum AAA diameter (Dmax).

<table>
<thead>
<tr>
<th>Observer type</th>
<th>Observations (n)</th>
<th>Mean paired difference (mm)</th>
<th>SD (mm)</th>
<th>Frequency of observations</th>
<th>≤2mm</th>
<th>≤5mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>540</td>
<td>-2.0</td>
<td>4.0</td>
<td>233 (43%) 423 (78%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Numbers in the parenthesis highlights the percentage for each category. All continuous data is expressed in millimetres unless otherwise stated. SD, standard deviation; Dmax, maximum AAA diameter.

Table 4. Variability between observers when selecting CT slices for Dmax measurements.

<table>
<thead>
<tr>
<th>Slice difference</th>
<th>Radiologist</th>
<th>Radiographer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27 (50%)</td>
<td>269 (42%)</td>
<td>296</td>
</tr>
<tr>
<td>1</td>
<td>22 (41%)</td>
<td>301 (46%)</td>
<td>323</td>
</tr>
<tr>
<td>2</td>
<td>5 (9%)</td>
<td>76 (12%)</td>
<td>81</td>
</tr>
<tr>
<td>3</td>
<td>0 (0%)</td>
<td>2 (0.3%)</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>648</td>
<td>702</td>
</tr>
</tbody>
</table>

Chi-squared analysis revealed no significance difference between groups (P=0.228).