Radiography (2010) **16**, 321—326

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/radi

A comparison of radiographers and radiologists in CT based measurements of abdominal aortic aneurysms

Andrew England*, Abigail Best, Charlotte Friend

Directorate of Medical Imaging and Radiotherapy, University of Liverpool, Johnston Building, Quadrangle, Brownlow Hill, Liverpool L69 3GB, United Kingdom

Received 27 October 2009; revised 4 February 2010; accepted 28 February 2010
Available online 21 March 2010

KEYWORDS
Observer variability;
Abdominal aortic aneurysm;
Computed tomography

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 measure maximum aneurysm diameter (Dmax) on ten CT scans. Each CT scan was presented randomly to each observer; four were duplicates testing intra-observer variability. All measurements were undertaken 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 ≤ 2 mm or less and 78% ≤ 5 mm. 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 encouraging. Variability exists for both professions, and in some instances may be clinically 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.

* Corresponding author. Tel.: +44 151 794 5123; fax: +44 151 794 5719.
E-mail address: a.england@liv.ac.uk (A. England).

1078-8174/$ - see front matter
doi:10.1016/j.radi.2010.02.008

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; AAAs are often asymptomatic and if left undetected they will continue to expand and may eventually rupture. Decisions about the treatment of AAAs are traditionally based upon the maximum cross-sectional diameter. If the AAA diameter is 5.5 cm or larger then intervention is generally deemed appropriate.[1] AAAs 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; if an AAA was measured to be significantly smaller than its actual size then the patient would be treated conservatively and exposed to a high risk of rupture. Conversely if the AAA is measured to be larger than its 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 difference between radiographers and radiologists in undertaking CT measurements of AAAs. 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[5] 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 abdomen and demonstrated an AAA with a variety of maximum diameters. CT acquisition comprised of 5 mm full-field axial CT images of the abdominal aorta following injection of iodinated contrast media with images reconstructed every 5 mm. For the purposes of this study images were assessed using the computer software Dicom Works v1.3.5 (DicomWorks, Lyon, France).

Observers (consultant vascular radiologists or qualified diagnostic radiographers) were recruited from two University Teaching Hospitals in the north-west of England. Each hospital was a vascular tertiary referral centre and undertook CT assessment of a minimum of 10 + AAAs patients per week. Each observer was given a basic eye test to confirm that eyesight was not a variable. The 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 (400 W 40 L). 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 to observers 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 series. Before the study commenced all duplicate CT scans had several initial slices removed so that the scan start position varied in order to discourage any observers from recognising that duplicates were present. All duplicate CT scans were presented at a point in the test other than directly before or after its identical pair. For all measurements room lighting was kept standard (ambient) and measurements were undertaken in a room which resembled a standard radiological reporting room. The local research ethics committee approved this study prior to its commencement.

Statistical analysis

To assess the intra-observer variability the difference in paired measurements was calculated for repeat measurements of the same CT scan by the same observer. These differences were summarised as mean and standard deviation values together with 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). Inferential statistical analysis was 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 consultant vascular radiologists and 15 radiographers (mean post qualification experience 10.4 years, range 1–20 years). 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 ± 3.5 mm compared with -0.18 ± 2.6 mm for radiographers. Although differences existed between the two observer types and was maintained when categorising the paired differences ($\leq$5 mm) this did not reach statistical significance. Full results of the

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

Observer type	Paired Measurements (*n*)	Mean paired difference (mm)	SD (mm)	Frequency of observations	
				≤2 mm	≤5 mm
Radiologists	16	−2.1	3.5	10 (63)	11 (69)
Radiographers	60	−0.18	2.6	36 (60)	56 (93)

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. Inferential analysis between groups using the student *t*-test demonstrated no statistically significant difference ($P = 0.054$).

intra-observer analysis for each observer type are shown in Table 1 and Figs. 1 and 2.

Inter-observer variability within each observer group was calculated in order to allow the comparison of measurements made by observers of the same discipline (Table 2, Figs. 3 and 4). Mean paired differences were equivocal for both disciplines (radiologists -0.09 ± 3.1 mm; radiographers 0.1 ± 5.0 mm) but radiologists had increased consistency when paired differences were categorised. For radiologists 67% of paired measurements were within 2 mm and 91% were within 5 mm compared to 42% and 80% for radiographers, these results were not statistically significant.

Direct comparisons of the consultant radiologists' measurements with radiographers are displayed in Table 3 and Fig. 5. Inter-observer variability was good with 43% of paired measurements within 2 mm and 78% within 5 mm. Finally, radiologists and radiographers were compared on the slice selection chosen 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 AAA is a well established objective criterion for selecting patients for treatment and when assessing the results following endovascular repair (EVAR). There are currently no reports comparing the variability of CT measurements of AAAs between groups of radiologists and radiographers. Our data demonstrates (Table 1) that intra-observer mean differences were lower -0.2 ± 2.6 mm for radiographers when compared to radiologists. This may suggest that radiographers were internally more reliable and accurately replicated the task when asked to do so. However, when reviewing the limits of agreement and scattering on Bland—Altman plots (Figs. 1 and 2) both observer types appeared to perform equally well. Inferential analysis confirmed that there was no statistically significant difference in the intra-observer variability between radiologists and radiographers. When comparing the proportion of paired measurements within 2 mm of less, both groups averaged around 60% of measurements within this criterion. A possible reason for the identified difference in intra-observer variability may be due to the lower numbers of observers in the radiologist group ($n = 4$) relative to the radiographer group ($n = 15$).

Inter-observer variability within each group was assessed, examining the mean paired difference and SD suggests that both groups performed equally well. Despite this, when analysing the number of measurements within 2 mm and 5 mm, radiologists had more pairs within each of these categories 67% and 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 (Fig. 3). Data presented in Table 3 confirms 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 ≤5 mm this level is generally considered to be the criterion for establishing an AAA size change.[7,8] Further review of Bland—Altman plots and limits of agreement suggests that

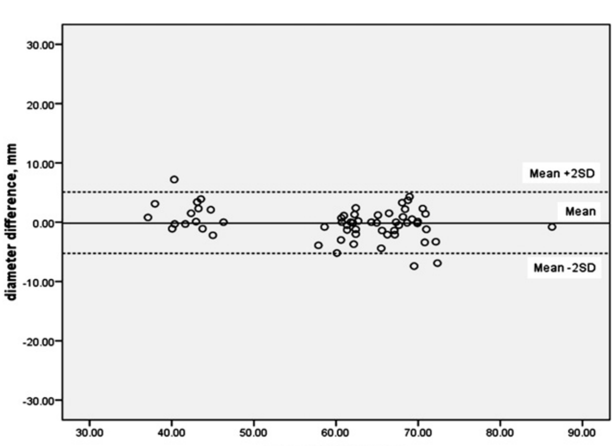

Figure 1 Bland—Altman plot demonstrating intra-observer variability between radiographers.

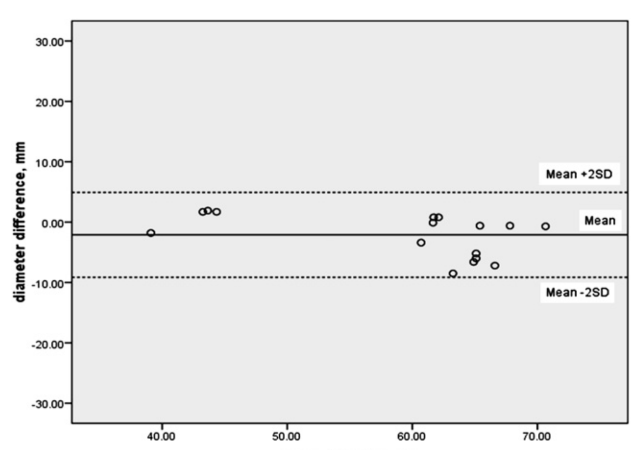

Figure 2 Bland—Altman plot demonstrating intra-observer variability between radiologists.

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

Observer type	Observations (n)	Mean paired difference (mm)	SD (mm)	Frequency of observations	
				≤2 mm	≤5 mm
Radiologists	54	−0.09	3.1	36 (67%)	49 (91%)
Radiographers	945	0.10	5.0	396 (42%)	736 (80%)

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. Inferential analysis between groups using the student t-test demonstrated no statistically significant difference ($P = 0.680$).

observer variability when assessed between groups is only marginally greater than within groups and between the same observer.

A study by Lederle et al.[7] compared local CT measurements with a central core laboratory. They found 17% of paired measurements were ≥5 mm which compares favourably to our study where 22% of paired measurements were ≥5 mm. What is not clear from the report by Lederle et al.,[7] is the training and experience of the core laboratory observers.

Jaakkola et al.[8] compared US and CT measurements of AAAs; they reported that 62% of paired CT measurements undertaken by radiologists were within ≤2 mm of each other. This compares well with our study where consultant radiologists produced 67% of paired measurements within ≤2 mm of each other. Reports investigating the variability between radiologists and non-radiologists are extremely limited. England et al.[4] assessed variability of aortic CT measurements by comparing a single radiologist with a single radiographer. There study highlighted a mean difference of 0.56 ± 3.2 mm between the two observers. Data from this study demonstrates an increased variability with mean differences of -2.0 ± 4.0 mm between the two disciplines. Results between the two studies cannot be directly compared, in the original study by England image analysis was undertaken on a 3D computer workstation and involved a single observer from each discipline. Despite this the difference between the two studies may be caused by the difference in numbers of observers between studies. Our report uses multiple observers from each discipline, there is some evidence within the literature advocating that variability may increase when using many observers[9] or may be lowered if using 3D imaging techniques.[10,12]

Several comments can be made about this study. Understanding the clinical significance of measurements is an important factor. With diameters in excess of 5.5 cm being universally accepted as the treatment size criterion arguments could be raised about the accuracy of measurements of clearly small or exceptionally large aneurysms. Measuring 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 may be explained by increases in the paired differences for small and large AAAs when reviewing the Bland–Altman plots (Fig. 5). Such differences were not present in our study and this may reflect the clearly defined measurement instructions given to each participant. It can therefore be deduced from this study that measurement variability was independent of AAA size. If clear and precise instructions resulted in lower variability then our finding supports statements by Cayne et al.[10] Cayne and colleagues argue that inter-observer variability cannot be eliminated but can be reduced by standardisation of measurement techniques.

Another possible problem with this study is the effect of the aorta bending within the abdomen. A tortuous aorta could lead to the aorta being imaged obliquely at its maximum extent and this may make the AAA much larger when assessed. It may be possible that more experienced observers e.g. radiologists may take this into account when

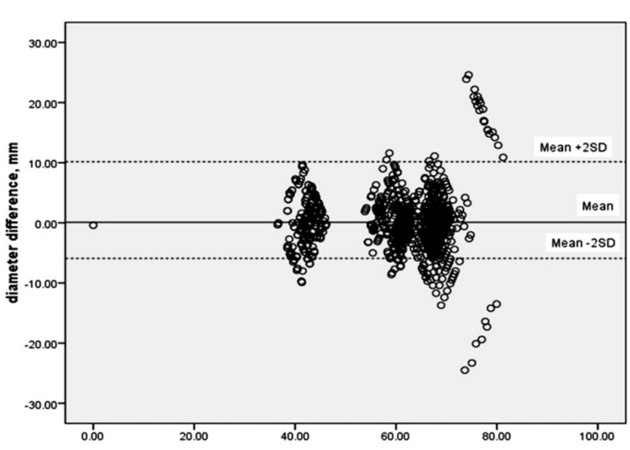

Figure 3 Bland–Altman plot demonstrating inter-observer variability between radiographers.

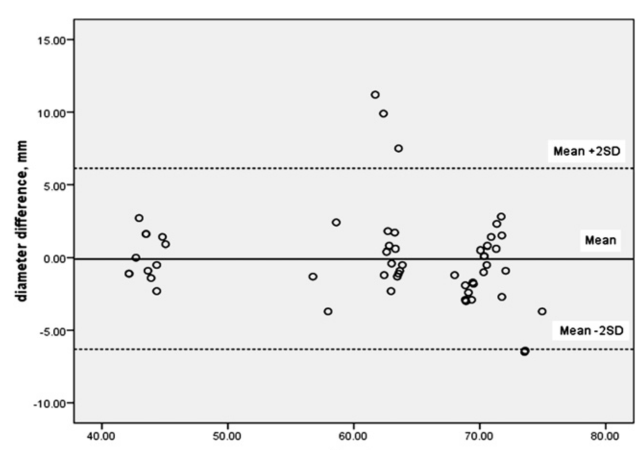

Figure 4 Bland–Altman plot demonstrating inter-observer variability between radiologists.

Table 3 Inter-observer variability between observer groups when measuring maximum AAA diameter (Dmax).

Observer type	Observations (n)	Mean paired difference (mm)	SD (mm)	Frequency of observations	
				$\leq$2 mm	$\leq$ 5mm
Combined	540	−2.0	4.0	233 (43%)	423 (78%)

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.

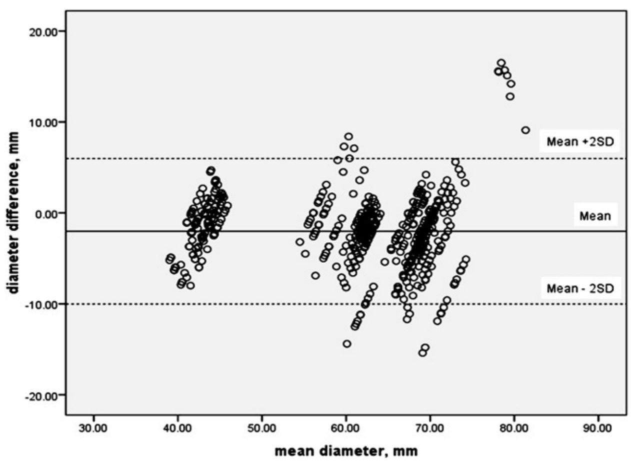

Figure 5 Bland—Altman plot demonstrating inter-observer variability between radiographers and radiologists.

measuring and thus this could affect variability. Future studies evaluating measurement variability on 3D workstations is warranted and may demonstrated significant improvements. Further studies in this area have been advocated by Broeders et al.[12] and by Cayne et al.[10] In addition to 3D workstations Cayne et al.[10] argues the need to consider the utility of computer-assisted measurements and other AAA growth variables e.g. volume. Arguing against this last point Abada et al.[11] in their study, concluded that maximum diameter measurements are sufficient for sizing AAAs and that volume measurements should only be used in select cases where diameters do not allow a clear classification. It is clear from all publications that adequate consideration of observer variability must be given when performing AAA measurements, especially when deciding treatment.

The measurement of maximum AAA diameter from CT scans generates two tasks 1) identification of the slice where the aorta is depicted at its maximum diameter and 2) selecting two points on an image which demonstrates the maximum dimensions of the AAA which forms the measurement. Variability is almost certainly due to both processes and will also be affected by external factors such as observer experience and visual contrast sensitivity. From our data (Table 4) it would appear that variability is less influenced by slice selection and more by the decisions surrounding plotting of the measurement points on the axial CT image. Discussions with study participants suggest 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 may be dependent on a person's visual ability to clearly identify junctions between low contrast structures. Further investigation in this area is planned and will assess the effect of visual contrast sensitivity on aortic CT measurement accuracy.

Role extension within radiographic practice is now well established, despite this there is a lack of evidence surrounding the accuracy of radiographers in undertaking measurements from medical images. There is encouraging agreement between radiologists and radiographers when performing aortic CT measurements. To understand whether this has any practical potential within the NHS 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 possible benefits to patients.

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

Slice difference	Radiologist	Radiographer	Total
0	27 (50%)	269 (42%)	296
1	22 (41%)	301 (46%)	323
2	5 (9%)	76 (12%)	81
3	0 (0%)	2 (0.3%)	2
Total	54	648	702

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

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, whether 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 latter is to be accepted then these measurements will require validation and clinical checking before prescribing 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.

References

1. Ballard DJ, Filardo G, Fowkes G, Powell JT. Surgery for small asymptomatic abdominal aortic aneurysms. *Cochrane Database Syst Rev* 2008;(4):CD001835.
2. Lederle FA, Johnson GR, Wilson SE, Ballard DJ, Jordan Jr WD, Blebea J, et al. Rupture rate of large abdominal aortic aneurysms in patients refusing or unfit for elective repair. *JAMA* 2002 Jun 12;**287**(22):2968—72.
3. Wanhainen A, Bergqvist D, Bjorck M. Measuring the abdominal aorta with ultrasonography and computed tomography — difference and variability. *Eur J Vasc Endovasc Surg* 2002 Nov;**24**(5):428—34.
4. England A, Butterfield JS, Ashleigh RJ. Observer variation in vascular CT measurements of the abdominal aorta. *Radiography* 2008 Nov;**14**(4):282—7.
5. http://ludens.cl/photo/montest.html.
6. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 Feb 8;**1**(8476):307—10.
7. Lederle FA, Wilson SE, Johnson GR, Reinke DB, Littooy FN, Acher CW, et al. Variability in measurement of abdominal aortic aneurysms. Abdominal Aortic Aneurysm Detection and Management Veterans Administration Cooperative Study Group. *J Vasc Surg* 1995 Jun;**21**(6):945—52.
8. Jaakkola P, Hippelainen M, Farin P, Rytkonen H, Kainulainen S, Partanen K. Interobserver variability in measuring the dimensions of the abdominal aorta: comparison of ultrasound and computed tomography. *Eur J Vasc Endovasc Surg* 1996 Aug;**12**(2):230—7.
9. Aarts NJ, Schurink GW, Schultze Kool LJ, Bode PJ, van Baalen JM, Hermans J, et al. Abdominal aortic aneurysm measurements for endovascular repair: intra- and interobserver variability of CT measurements. *Eur J Vasc Endovasc Surg* 1999 Dec;**18**(6):475—80.
10. Cayne NS, Veith FJ, Lipsitz EC, Ohki T, Mehta M, Gargiulo N, et al. Variability of maximal aortic aneurysm diameter measurements on CT scan: significance and methods to minimize. *J Vasc Surg* 2004 Apr;**39**(4):811—5.
11. Abada HT, Sapoval MR, Paul JF, de Maertelaer V, Mousseaux E, Gaux JC. Aneurysmal sizing after endovascular repair in patients with abdominal aortic aneurysm: interobserver variability of various measurement protocols and its clinical relevance. *Eur Radiol* 2003 Dec;**13**(12):2699—704.
12. Broeders IA, Blankensteijn JD, Gvakharia A, May J, Bell PR, Swedenborg J, et al. The efficacy of transfemoral endovascular aneurysm management: a study on size changes of the abdominal aorta during mid-term follow-up. *Eur J Vasc Endovasc Surg* 1997 Aug;**14**(2):84—90.