Reliability of aortic aneurysm screening measurements

Rachel Weston Smith¹, Tony Evans¹ and Stephen Wolstenhulme²

¹Department of Medical Physics, University of Leeds, Leeds, UK

²School of Healthcare, University of Leeds, Leeds, UK

Corresponding author: Rachel Weston Smith. Email: um09rws@leeds.ac.uk

Abstract

The aim of this feasibility study was to assess the impact that image processing of abdominal aortic ultrasound (US) images had on the intra-observer reliability of the diameter measurement. The study compared variability between inner-to-inner (ITI), outer-to-outer (OTO) and outer-to-inner (OTI) wall diameter measurements and their resilience to image processing. Three US images of transverse abdominal aortas were manipulated in 13 different ways using functions from Image J software (National Institutes of Health, Bethesda, MD, USA). Blinded measurements were performed of the aortic diameters from each image; this was repeated for ITI, OTO and OTI. Profiles of each image were produced and sets of rules developed to provide detailed instructions as to where, on the profile, the callipers should be placed to correspond with the actual image. The reliability of the diameter measurements compared to the original diameter measurement was least affected by adjusting the brightness and contrast of the US images (better than ± 1.5 mm). Using the functions 'Sharpen' and 'Find Edges' created the largest difference (up to -5 mm). The ITI measurements had the widest spread of variability, whereas the OTI measurements proved to be the most repeatable and resilient to image processing. This study suggests the precision of the measurements can be kept within satisfactory levels even after image manipulation. It also showed the most reliable measuring guideline was OTI, in contrast to the guideline currently used by the NHS Abdominal Aortic Aneurysm Screening Programme. Further research is needed to transfer the findings into the clinical setting of the National Screening Programme to increase its reliability.

Keywords: Ultrasound, abdominal aortic aneurysm, diameter, measurement

Ultrasound 2014; **22**: 80–90. DOI: [10.1177/1742271X13513980](https://doi.org/10.1177/1742271X13513980)

Introduction

Internationally, mortality associated with abdominal aortic aneurysms (AAA) is of great concern; in England and Wales alone, ruptured AAAs cause more than 6800 deaths annually.¹ In response to this, the NHS Abdominal Aortic Aneurysm Screening Programme (NAAASP) was established in 2009 with the aim of identifying at-risk aneurysms and facilitating interventions to prevent spontaneous rupturing of abdominal aortas.² Errors in the ultrasound (US) measurements may reduce the effectiveness of the programme by miscategorising patients and therefore it is right to consider all factors which might lead to inaccuracies in the measurement. The reliability of US measurements in this context depends upon three key factors; the machine, the measurement guidelines and the operator. The measurement guidelines are concerned with ensuring that a true transverse section is selected, although interestingly there is no guidance as to the phase of the cardiac cycle which is to be used.³ In addition, the guidelines prescribe the positioning required for the machine callipers. Machine variations

can be minimised by ensuring that all equipment in the screening programme is calibrated to the same assumed speed of sound (1540 m s^{-1}). However, the accuracy and precision of the measurement may also vary between machines because of variations in spatial resolution and, in the case of axial measurement, differences in the demodulation algorithms used. There is evidence from nuchal translucency screening and elsewhere that the use of harmonic imaging can lead to variations in measurements obtained.⁴ All other things being equal, it seems logical to select measurement guidelines in each case, which are as resilient as possible when different signal processing algorithms are used.

For the NAAASP, the US examination is carried out in a community setting by operators whose training is less intensive than that of general sonographers. In similar situations, such as bladder volume measurement,⁵ the measurement has been automated to reduce operator dependence further. This approach has been suggested for use in automation of measurements of the aortic diameter.⁶ It would

therefore be of interest to determine the optimum methodology, which might be used in such cases.

In this study, we have investigated the impact which a range of different image processing options have on the reliability of the measurement. We have identified the optimal choice of image processing, which is least likely to lead to variation between machines, signal processing, power output, frequency, display settings and room lighting. It might be assumed that such a choice will also be the most appropriate one when automation of the measurement is considered. The second potential source of variability is the calliper positioning on the antero-posterior (A-P) walls of the aorta, as prescribed in the measurement guidelines.^{7,8} The NAAASP guidelines specify the use of the inner to inner wall diameter (ITI) as reported in the Multicentre Aneurysm Screening Study.⁹ The inner wall is not the only feasible landmark and not necessarily the most reliable one.⁷ The outer to outer wall (OTO) and outer to inner wall diameters (OTI) can also be selected.⁸ The ITI and OTO measuring methods have been evaluated by several studies, but with conflicting conclusions on which method produced the most reproducible results.^{10–12} There is little research into the validity of the OTI method; the only paper found a repeatability similar to the OTO method.³ A-P measurements are axial, so will be influenced by signal processing algorithms. Transverse diameter measurement uncertainties are dominated by beam width and focussing limitations and may lead to poorer accuracy.⁷

Finally, reliability is a function of the operator. Measurements taken can vary between ultrasound practitioners (inter-observer repeatability) and between multiple measurements taken by one person (intra-observer repeatability).^{7,8,10,13} One possible method of reducing this human error may be the introduction of automation in the measurement. To implement this, a clear set of rules needs to be identified to determine A-P calliper positioning. Software to carry out automation would be expected to be based on the use of one or more profiles (similar to an A-scan) extracted from the image, on which calliper location is superimposed.

The aim of this feasibility study was to evaluate the reliability of ITI, OTO and OTI measurements of abdominal aortic diameter when subjected to image processing manipulation. A secondary aim was to determine the optimal method of using profiles for aortic diameter measurements as an intermediate step towards automation.

Methodology

Image selection

Images were selected from a set of 900 US images acquired from nine different machines in a single centre, as part of a larger study into equipment specification funded by the NAAASP. The set of 900 contained 450 longitudinal plane and 450 transverse plane images. For the transverse images, a subset of 27 was selected, consisting of three images of different patients from each of the nine machines. The images were selected by identifying those at which the aorta was positioned at a depth of approximately 10 cm; the mean value in the set. The depth was standardised to reduce the variability which would be introduced by attenuation, which increases with depth. The 27 images were then used in an image evaluation study conducted during the British Medical Ultrasound Society (BMUS) meeting (Telford, December 2012).¹⁴ Emergency medicine physicians, obstetricians, physicists, radiologists, screening technicians, sonographers and vascular technologists participated in a quality scoring assessment. The conclusion was that two images emerged as having the 'best' and 'worst' quality. Coincidentally, both images were obtained using the same machine; a third image was added from this machine. In this way, the three images selected, labelled P1 (Best), P2 (Intermediate) and P3 (Worst) (Figures 1(a) to (c)) came from the same machine, operated by one individual, involving three different patients. By selecting best, worst and intermediate quality images, a broad range of tissue types were covered. No patient identifiable data was included in the images, which were acquired as part of the routine patient examination and so it was not necessary to submit the study for formal ethical approval.

Image manipulation

Each of the three images was manipulated in 13 different ways using image processing options available from Image J software (National Institutes of Health, Bethesda, MD, USA) (Figure 2).¹⁵ The specific manipulations were chosen because they could be implemented readily and had obvious effects on the images. Many other image processes could have been applied, but due to time constraints, these were not explored. The processes used were Brightness, Contrast, 'Contrast Enhancement', 'Find

Figure 1 The three images selected for the study: (a) P1 – best quality, (b) P2 – intermediate quality and (c) P3 – worst quality

Figure 2 Original P1 Image plus all 13 image processing manipulations. These are: (a) Original Image, (b) -50% Brightness of original image, (c) +50% Brightness of original image, (d) +100% Brightness of original image, (e) -50% Contrast of original image, (f) +50% Contrast of original image, (g) Contrast Enhancement on, (h) Find Edges on, (i) -50% Level of original image, (j) +50% Level of original image, (k) -100% Level of original image, (l) Sharpen on, (m) -50% Window of original image, (n) +50% Window of original image

Edges', 'Level', 'Sharpen' and 'Window'. Four of the functions, Brightness, Contrast, Level and Window, had an adjustable scale and so were evaluated at plus and minus 50% of the original setting. In addition, 100% settings of Brightness and Level were used, but this was not practical for Contrast and Window as the images either became too dark or too bright to utilise. The other three processes were available on an on/off basis, so were only implemented once on each image. This created 42 US images; 39 manipulated and the three original images (Figure 2a).

Ultrasound image aortic diameter measurements

The images were placed in a random order using the randomise function in Microsoft Excel. One observer, who did not routinely undertake US examinations (RWS), carried out blinded measurements of the aortic diameters over two consecutive days. On the first day, the 42 images were measured twice, and on the second day, the same 42 images were measured twice again. The ITI, OTO and OTI A-P abdominal aortic diameters were measured and this was repeated four times for each image process and each diameter measuring method. The measurements were performed using Image J and undertaken in a room with reduced lighting to optimize the viewing capability and replicate the clinical environment.¹⁶

The diameter results, initially measured in pixels, were calibrated into millimetres. Depth markings on the US images were removed as part of the anonymisation procedure, so another US image from the same machine with depth markings was used for calibration.

Averages of the measurements were calculated and the difference between the diameter of the original images and the diameters of the manipulated images were determined. From this, GraphPad Prism (GraphPad Software, La Jolla California USA)¹⁷ was used to plot the variation found for each method of measuring the diameter (Figures 4(a) to (c)).

Aortic diameter profiles

A variety of sets of rules were used to define where on the profiles the callipers should be placed to correspond with the calliper placement on the actual US image. The sets of rules were arbitrary and could have been formulated in many different ways. Once they were established, profiles of each aorta from both the original and manipulated images were constructed using Image J. The ITI, OTO, OTI and baseline diameters were measured for each profile.

To obtain the profiles, two points were selected on the US image, the first proximal to the anterior wall of the aorta and the second distal to the posterior wall. A vertical line was then drawn between the two points before using the function 'profile plot' on Image J (Figure 3a). The first step, after gaining the profile, was to identify the anterior and posterior walls of the aorta (shaded area on Figure 3c). It was then important to observe the lumen of the aorta and decide the baseline D (Figure 3c). The lumen, which is hypoechoic on the US image, appears undulating on the profile plot, suggesting the lumen contains artefacts and noise. This reduces as the depth increases within the

lumen. The OTO diameter A (Figure 3c) was readily defined in most cases. The reference points were the bottom of the upstroke peak corresponding to the anterior wall and the bottom of the down stroke peak corresponding to the posterior wall. The diameter was measured at the same grey level. Therefore, if the two reference points did not lie along the same level, a horizontal line was drawn from the first reference point until it encountered the vertical line that passed through the second reference point (this is not illustrated in Figure 3 where the OTO grey levels are the same). Finding the reference points for the ITI diameter, C (Figure 3c) was problematic. Initially, the baseline of the lumen was defined; a subjective decision was made as to where the signal from the inner walls of the aorta changed to the level in the lumen, D (Figure 3c). The two ends of the baseline then allowed the thickness of the anterior and posterior walls to be measured (black arrow on Figure 3c). However, this does lead to equal values for the thickness of both walls. Pragmatically, we selected the proximal wall as the more reliable and chose a distance δ equal to one-quarter of its value. Starting at the baseline, a point was selected at a distance δ and a vertical line was then drawn from this point to its intersection. A horizontal line was then drawn from the point of intersection to represent the ITI value. The OTI, B (Figure 3c) was measured using the first reference point of the OTO and the second reference point of the ITI diameters.

A Shapiro-Wilk test demonstrated that the profile data was not normally distributed ($p > 0.05$), so a Spearman's test was used to calculate the correlation between the profile plot measurements and the diameters taken from the US images using SPSS Software (SPSS Inc. Chicago, USA). These results were statistically significant ($p < 0.05$).

Results

Ultrasound image aortic diameter measurements

The repeatability of the four values for each image was different for each of the three guidelines, ITI, OTO and OTI. The mean difference between the four measurements for the ITI was 1.53 (range 0.45–5.83 mm), for the OTO was 1.54 (range 0.02–8.53 mm) and for OTI was 0.70 (range 0.03–2.70 mm).

For each of the 13 image processing manipulations, the variation between measured values for all three guidelines and all three images is shown in Table 1.

Figure 4(a) to (c) shows the variations for the each measurement guideline. This shows that no specific image processing function gave the most reliable results. A small number of the functions used caused less deviation from the original image diameter than others. Manipulating the Brightness and Contrast of the US image affected the reliability least ($< \pm 1.5$ mm). Using the functions 'Sharpen' and 'Find Edges' created the largest difference, up to -5.0 mm. The adjustable functions produced less variation than the on/off functions. For P2 and P3 images, the majority of the diameter measurements showed an increase compared to the original image. In contrast, the P1 images showed a decrease in diameter.

Figure 3 An abdominal aortic ultrasound image and a profile of the corresponding diameter. This shows (a) the original P1 image, (b) is an expanded version and (c) the profile plot. The lines in (b) and (c) represent the OTO (A), OTI (B) and ITI (C) measuring methods and the baseline of the aorta (D). The distance is given in pixels and mm following calibration

Table 1 The mean difference between the manipulated images and profiles from the original images

Image processes	Mean difference from original image (±mm)	Mean difference between profiles and original image (±mm)
–50% Brightness from original image	0.68	0.52
+50% Brightness from original image	0.59	0.49
+100% Brightness from original image	0.75	0.46
–50% Contrast from original image	0.29	0.39
+50% Contrast from original image	0.39	0.32
Contrast enhancement	0.62	0.47
Find edges	2.99	2.01
–50% Level from original image	0.53	0.49
+50% Level from original image	0.59	0.50
–100% Level from original image	1.09	0.67
Sharpen	0.77	0.59
–50% Window from original image	0.52	0.39
+50% Window from original image	0.63	0.38

The OTI diameter measurements for all the images showed the least deviation from the measurement of the original image; 90% of the deviations were between –0.50 mm and + 0.75 mm, and 95% between –2.00 and

+2.00 mm (Figure 4a). The ITI diameters showed that 90% of measurement variations were between –2.0 mm and +1.50 mm and 95% were between –2.74 and +2.24 mm (Figure 4b). Also 90% of OTO variations were between

Figure 4 Plots showing the deviation of (a) OTI, (b) ITI and (c) OTO wall diameter measurements of the manipulated images from the original image reference measurement

−0.50 mm to +1.50 mm and 95% were between −2.4 and +2.8 mm (Figure 4c). The relatively large differences seen between the two cut-off values (90% and 95%) were associated with very large values in a small number of outliers.

The P2 image (Figure 1b), which was classed as having a mid-range quality, showed the widest variation for all three measurement guidelines; the deviation from the original measurement was between +1.87 mm and −5.12 mm.

Figure 4 (Continued)

The image with the best quality, P1 (Figure 1a), had a similar range (+1.48 mm to -4.86 mm). The image with the worst quality, P3 (Figure 1c), had the least deviation from the original (range +1.81 mm to -1.57 mm).

Aortic diameter profiles

The profile measurements gave similar diameters to those measured from the US images. All the correlations were strongly positive and none less than 0.80 (Table 2).

Figure 4 (Continued)

The OTO diameter on the US image, 'A' on the profile (Figure 3), displayed the strongest correlation (0.92) and that for the OTI measurement, 'B' on the profile (Figure 3) was also similar (0.91). However the ITI, 'C', diameter correlation was less positive (0.84).

The deviation using the ITI rule had the largest variation, (-3.43 mm to $+2.30$ mm). The OTO diameter values showed a significant deviation from that of the B-scan images (2.07 mm to $+1.95$ mm). The OTI diameter measurement proved to be the most robust rule (-1.12 mm to $+0.68$ mm differences).

Evaluating the difference between the measurements from the US images and the profiles enabled an alternative assessment of the various image processes. Each function used to manipulate the images changed the individual profile plots. The 'Find Edges' function produced the most varied diameter measurement for all of the different diameters, especially for the images P1 and P3. This is consistent with the finding from the US images. Other functions that were less successful were Sharpen and increasing the Brightness and Level to maximum. Altering the Brightness, Contrast and Level all created profile plots that had comparable diameter measurements to the US images; the lengths were within ± 0.09 mm of the US measurements.

Discussion

Image manipulation

This study has shown that image processing can increase and decrease the reliability of aortic diameter measurements (Figures 4(a) to (c)). The functions, Brightness and Contrast do not affect the reliability by a significant amount, less than ± 1.5 mm. This was below the acceptable observer error recommended by NAAASP (± 5 mm).¹⁸ However, the binary functions, Sharpen and Find Edges, altered the aortic diameters by an unacceptable amount, up to -5.0 mm. These findings suggest that certain image processes help to improve the quality of the US image by increasing their visualisation and can do this without affecting the reliability of the measurements taken from the image.

Ultrasound image aortic diameter measurements

The most significant outcome from this research was the striking difference between the reliability of the ITI, OTO and OTI diameters. The OTI measuring method was the most accurate and resilient to image processing (Figure 4a), suggesting that these landmarks may be the most robust for calliper placement. Wanhainen concluded that this method of leading edge to leading edge should theoretically produce the most reliable results.¹⁹ It uses the vessel walls with the most distinct US reflection, where sound travels through boundaries from an echo-lucent to an echo-dense layer, and therefore are the easiest to locate the position of the callipers. Sweden already employs this method into their national screening programme;²⁰ the outcomes may be interesting to compare.

Altering the current NAAASP guidelines may affect the sensitivity and specificity of the programme. It seems probable that the actual number of men whose classification might be altered by making this sort of alteration is relatively small. However, if the results of this study are confirmed, then an improvement in the reliability of the programme could be achieved at almost no cost.

Aortic diameter profiles

Although ultrasonography is a consistent screening tool, reliably providing comparable aortic diameters with different US machines in separate centres across the country is

Table 2 The Spearman correlation coefficients between all the diameter measuring methods from the US images and the equivalent profile plots

Spearman's correlation coefficients	Profile plot measurements				
	Ultrasound image measurements	A (OTO)	B (OTI)	C (ITI)	D (Baseline)
OTO		0.92	0.91	0.87	0.82
OTI		0.92	0.91	0.89	0.84
ITI		0.85	0.83	0.84	0.87

difficult. Continuous review and research are needed to help improve the screening programme. The final part of the study showed that measuring the aortic diameter from a profile was positively correlated with measuring the aortic diameter from an US image, especially for the OTO and OTI diameters (Table 2). This may be evidence that automated measuring systems could be used to measure aortic diameters in the NAAASP. Furthermore, in vivo studies using signal processing embedded in ultrasound machines and a larger sample size would be needed to reach a more concrete conclusion.

A major issue facing the NAAASP is the minimal training and expertise of the employed screening technicians (STs). Hartshorne *et al.* found significant variability between specific US sonographers and STs when measuring AAA.¹⁰ If an automated system could be produced where the operating system measured the diameter independently, a significant source of error may be eradicated. Abbas *et al.* looked into this idea by testing the Aorta Scan AMI 9700 (Verathon, Bothell, WA, USA) machine which automatically measures the maximum diameter of the aorta.²¹ They concluded that it has the potential to move the screening programme into the community without the need for trained US operators; however, it still needs technical improvement before it can replace STs.

Limitations

This study can only account for intra-observer variability, as the US measurements were taken by one person who has had no NAAASP accredited US training. Only static US images were used in the study, therefore the conclusions drawn may not be comparable with the screening programme. The three images utilised all exhibited aortas with no or minimal aneurysm; variability is greater for larger aortic diameters.²² Therefore, the results found may underestimate the full scale of the problem. This suggests that the precision and reliability for aneurysmal aortic measurements may be greater, although caution is indicated here since Beales *et al.* have concluded that results from different studies are contradictory.⁷ Implementation of automated diameter measuring systems may not transfer reliably between different US machines.

There was a limitation arising from the choice of statistical test. Correlation coefficients between two data sets such as these have a high predictive value because although

two different measuring techniques were used, the same structure was measured.²³

Overall, this study can be described as a feasibility study in this area. Using a low number of images from only one machine and with a single observer is a useful method of limiting the number of variables. Nonetheless, the results obtained are consistent with findings in similar situations and many of the variations found are significant.

The measurement of aortic diameter is subject to many sources of error. These include patient variables such as aortic depth, vessel tortuosity and obesity, which degrade the quality of the image. The choice of phase in the cardiac cycle and breathing movement will also influence the value obtained, as will vessel obliquity if not carefully taken into account. Further complications will arise if intra-luminal thrombus is present. It might therefore seem that optimisation of calliper positioning would be insignificant. We have not sought to make a relative estimate of the contribution which these various errors will make, but it would seem wrong to not optimise those things which can be readily addressed.

Recommendations for future research

To develop the project, the number of images should be increased and taken from a variety of US machines, with a range of normal aortas to a variety of small and large aneurysms. The participation of several trained STs would test both intra- and inter-observer variability. Future research should explore image processing further and apply it to a clinical setting as well as investigate the feasibility of implementing an automated measuring system throughout the screening programme. Examining the possibility of using US radiofrequency scans could be a variation on an automated measuring device. Finally, continuous research into the most reliable aortic measuring method is fundamental to the NAAASP achieving precise and reliable guidelines.

Conclusion

Overall, the accuracy of the diameter measuring method is of an acceptable level in current practice and the overall impact of the inter- and intra-observer variability is minimal. However, ongoing evaluation and audit of the national screening programme is needed to ensure that the public is receiving the best preventative care. This study suggests that the accuracy of the measurements can be kept within satisfactory levels even after major manipulation of the US image. Brightness and Contrast have been identified as potential manipulations that could improve the visualisation of the aorta, thereby increasing the accuracy of the calliper placement without affecting the repeatability. The NAAASP, like all screening programmes, is subject to constant evaluation and this allows for ideas for development and progression. The application of some image processing and the use of automated profile measurement are just two ways in which reliability of the screening programme could be improved. The principal conclusion from this study is that the current measuring guideline of the NAAASP may not be the most reliable

and that the OTI wall diameter measurement is consistently more reliable. This should be explored further to avoid unnecessary miscategorisation which may impact on triage decision-making.

DECLARATIONS

Competing interests: The authors have no conflicts of interest to declare.

Funding: None.

Ethical approval: Not applicable.

REFERENCES

1. UK National Screening Committee. *Structured review for the UK National Screening Committee Appraising the Viability, Effectiveness and Appropriateness of an Abdominal Aortic Aneurysm Screening Programme*. See <http://aaa.screening.nhs.uk/aboutus> (last checked 2 March 2013)
2. NHS Abdominal Aortic Aneurysm Screening Programme. *Abdominal Aortic Aneurysms Statistics*. See <http://aaa.screening.nhs.uk/aaastatistics> (last checked 5 March 2013)
3. Bredahl K, Eldrup N, Meyer C, et al. Reproducibility of ECG-gated ultrasound diameter assessment of small abdominal aortic aneurysms. *Eur J Vasc Endovasc Surg* 2013;**45**:235–40
4. Evans JA, Watson A, Coulson KA, et al. OC77: How accurate are NT measurements? *Ultrasound Obstet Gynecol* 2007;**30**:390–1
5. Saint S, Kowalski CP, Kaufman CA, et al. Preventing hospital-acquired urinary tract infection in the United States: A national study. *Clin Infect Dis* 2008;**46**:243–50
6. Reardon RF, Cook T, Plummer D. Abdominal aortic aneurysm. In: Ma OJ, Mateer JR, Blaivas M (eds). *Emergency Ultrasound*, 2nd edn. New York: Mc Graw-Hill, 2008:149–68
7. Beales L, Wolstenhulme S, Evans JA, et al. Reproducibility of ultrasound measurement of the abdominal aorta. *Brit J Surg* 2011;**98**:1517–25
8. Long A, Rouet L, Lindholt JS, et al. Measuring the maximum diameter of native abdominal aortic aneurysms: Review and critical analysis. *Eur J Vasc Endovasc Surg* 2012;**43**:515–24
9. Aneurysm Screening Study Group Multicentre. Multicentre aneurysm screening study (MASS): Cost effectiveness analysis of screening for abdominal aortic aneurysms based on four year results from randomised controlled trial. *BMJ* 2002;**325**:1135
10. Hartshorne TC, McCollum CN, Earnshaw JJ, et al. Ultrasound Measurement of aortic diameter in a National Screening Programme. *Eur J Vasc Endovasc Surg* 2011;**42**:195–9
11. Thapar A, Cheal D, Hopkins T, et al. Internal or external wall diameter for abdominal aortic aneurysm screening? *Ann R Coll Surg Engl* 2010;**92**:503–5
12. Sensier Y, Akil Y, Khodabakhsh P, et al. The impact of using inner-to-inner wall diameter measurement of abdominal aortic aneurysm. *Ultrasound* 2013;**21**:8–11
13. Jim J and Thompson RW. Clinical features and diagnosis of abdominal aortic aneurysm, UpToDate November 2012. See <http://www.uptodate.com/contents/clinical-features-and-diagnosis-of-abdominal-aortic-aneurysm> (last checked 20 February 2013)
14. Keeble C, Wolstenhulme S, Davies AG, et al. Is there agreement on what makes a good ultrasound image? *Ultrasound* 2013;**21**:118–23
15. Image J. See <http://rsbweb.nih.gov/ij/> (last checked 7 January 2013)
16. Moore SC, Munnings CR, Brettell DS, et al. Assessment of ultrasound monitor image display performance. *Ultrasound Med Biol* 2011;**37**:971–9
17. GraphPad Prism Software. See <http://www.graphpad.com/> (last checked 18 February 2013)
18. NHS Abdominal Aortic Aneurysm Screening Programme. *Annual Report 2009–2010*. See http://aaa.screening.nhs.uk/annual_report (last checked 16 March 2013)

19. Wanhainen A. Comments regarding 'Ultrasound Measurement of Aortic Diameter in a National Screening Programme'. *Eur J Vasc Endovasc Surg* 2011;**42**:2
20. NHS Abdominal Aortic Aneurysm Screening Programme. *International research*. See <http://aaa.screening.nhs.uk/international-research> (last checked 6 May 2013)
21. Abbas A, Smith A, Cecelja M, et al. Assessment of the accuracy of Aortascan for detection of abdominal aortic aneurysm (AAA). *Eur J Vasc Surg* 2012;**43**:167–70
22. Jaakkola P, Hippelainen M, Farin P, et al. Interobserver variability in measuring the dimensions of the abdominal aorta: Comparison of ultrasound and computed tomography. *Eur J Vasc Endovasc Surg* 1996;**12**:230–7
23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986;**327**:307–10