Reliability of Semiautomatic Centerline Analysis versus Manual Aortic Measurement Techniques for TEVAR among Non-experts

F. Rengier a,b,*, T.F. Weber a,b, S. Partovi a, M. Müller-Eschn er a,b, D. Böckler c, H.-U. Kauczor a, H. von Tengg-Kobligk a,b

a University Hospital Heidelberg, Department of Diagnostic and Interventional Radiology, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany
b German Cancer Research Center (dkfz) Heidelberg, Department of Radiology E010, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
c University Hospital Heidelberg, Department of Vascular and Endovascular Surgery, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany

Submitted 20 January 2011; accepted 8 April 2011
Available online 13 May 2011

Keywords: Thoracic aorta; Computed tomography; Three-dimensional imaging; Computer-assisted image analysis; Endovascular

Abstract

Objectives: The study aimed to test whether reliability and inter-observer variability of preoperative measurements for thoracic endovascular aortic repair (TEVAR) among non-experts are improved by semiautomatic centerline analysis compared with manual assessment.

Methods: Preoperative computed tomography (CT) angiographies of 30 patients with thoracic aortic disease (mean age 66.8 ± 11.6 years, 23 men) were retrospectively analysed in randomised order by one blinded vascular expert (reference standard) and three blinded non-expert readers. Aortic diameters were measured at four positions relevant to TEVAR using three measurement techniques (manual axial slices, manual multiplanar reformations (MPRs) and semiautomatic centerline analysis). Length measurements were performed using centerline analysis. Reliability was calculated as absolute measurement deviation (AMD) from reference standard and inter-observer variability as coefficient of variance (CV) among non-expert readers.

Results: For axial, MPR and centerline techniques, mean AMD was 7.3 ± 7.7%, 6.7 ± 4.5% and 4.7 ± 4.8% and mean CV was 5.2 ± 4.2%, 5.8 ± 4.8% and 3.9 ± 5.4%. Both AMD and CV were significantly lower for centerline analysis compared with axial technique (p = 0.001/0.042) and MPR (p = 0.009/0.003). AMD and CV for length measurements by centerline analysis were 3.2 ± 2.8% and 2.6 ± 2.4%, respectively. Centerline analysis was significantly faster than MPR (p < 0.001).

* Corresponding author. F. Rengier, University Hospital Heidelberg, Department of Diagnostic and Interventional Radiology, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany. Tel.: +49 (0) 6221 56 6410; fax: +49 (0) 6221 56 5730.
E-mail address: fabian.rengier@web.de (F. Rengier).

1078-5884/$36 © 2011 European Society for Vascular Surgery. Published by Elsevier Ltd. All rights reserved.
Thoracic endovascular aortic repair (TEVAR) has emerged as an accepted treatment option for thoracic aortic aneurysms and penetrating atherosclerotic ulcers. Precise preoperative assessment of the lesion's dimensions as well as of proximal and distal landing zones is necessary to select the appropriate endovascular stent-graft size and type. Inappropriate stent-graft sizing may cause development of postoperative complications, including endoleaks, aortic-wall trauma, aneurysm neck dilation and stent-graft collapse or dislocation. Hence, reliable measurements may reduce the incidence of postoperative complications which currently is up to 40%.

Computed tomography angiography (CTA) is the preferred imaging modality prior to thoracic and abdominal endovascular aortic repair because it offers various post-processing and measurement techniques. Previous studies on preoperative measurements focussed on the abdominal aorta and abdominal endovascular aortic repair. In this context, measurements were traditionally performed manually on axial CTA source data. However, this measurement technique suffers from substantial intra- and inter-observer variability. In addition, in contrast to the abdominal aorta, the thoracic aorta exhibits significant bending of the aortic arch and the descending thoracic aorta. Consequently, strictly axial measurements may not represent the true dimensions of the thoracic aorta because the axial plane can significantly differ from the true aortic cross section perpendicular to the vessel course.

Therefore, double oblique multiplanar reformations (MPRs), as arbitrarily adjusted planes perpendicular to the course of the aorta, have been suggested to be more accurate for the thoracic aorta. Semi-automatic centerline analysis might also improve measurement reliability. Centerline analysis is a post-processing technique that automatically generates MPRs perpendicular to vessel course and segments the vessel lumen after computational detection of the geometric vessel centre. It has been shown to reduce measurement variability and time in the abdominal aorta, but has not yet been evaluated with regard to TEVAR. The accuracy of automatic computations may be significantly influenced by the bending of the thoracic aorta. In summary, the reliability of different measurement techniques comprising use of axial source data, MPR and centerline analysis has not yet been investigated for the thoracic aorta, to our knowledge.

The purpose of this study was to test our hypothesis that reliability and inter-observer variability of diameter measurements for TEVAR among non-experts are improved by means of semiautomatic centerline analysis compared with manual assessment on axial slices and MPR. A further objective was to investigate differences between the measurement techniques, regarding time needed and training effect.

**Conclusions:** Semiautomatic centerline analysis provides the most reliable and least variable diameter and length measurements among non-experts in candidates for TEVAR.

**Materials and Methods**

**Patients**

The study was approved by the institutional review board and conducted according to its ethical guidelines. Written informed consent for investigations was obtained from all patients. Inclusion criteria for this retrospective study from January 2004 to April 2008 were: first, diagnosed with thoracic aortic aneurysm or penetrating atherosclerotic ulcer of the thoracic aorta; second, scheduled for elective TEVAR; and third, CT performed at our institution for preoperative planning. A total of 47 patients fulfilled all inclusion criteria. As many as 17 patients were secondarily excluded due to the following criteria: lesion not located in the descending thoracic aorta (n = 12) or second lesion in the descending thoracic aorta (n = 5). The latter exclusion criterion was defined to allow for standardised and thereby comparable measurements in all patients. The included 30 patients (seven women and 23 men) had a mean age of 66.8 ± 11.6 years (age range, 30–87 years).

**Image acquisition**

All patients underwent CTA examinations using two clinical multislice CT scanners, 17 patients on scanner A (Aquilion-16, Toshiba Medical Systems, Tokyo, Japan) and 13 patients on scanner B (Volume Zoom, Siemens Medical Systems, Erlangen, Germany). For scanner A, scan and reconstruction parameters were as follows: tube voltage 120 kV, tube current time product 120 mAs, reconstructed slice thickness 1.00 mm, reconstruction increment 0.80 mm, pixel spacing 0.6–0.75 mm and 90 ml contrast medium (iomepren with 400 mg iodine per ml, Imeron 400, Bracco Diagnostics, Princeton, NJ, USA) with 40 ml saline chaser. For scanner B, scan and reconstruction parameters were as follows: 120 kV, 120 mAs, reconstructed slice thickness 3.00 mm, reconstruction increment 3.00 mm, pixel spacing 0.6–0.75 mm and 120 ml contrast medium (iopromide with 370 mg iodine per ml, Ultravist 370, Bayer Health Care, Berlin, Germany) with 40 ml saline chaser.

**Image data preparation**

A research assistant, who was not involved in the image analysis, prepared the data for blinded analysis as follows. First, the 30 patients were randomised into two reading sessions of 15 patients. Second, three measurement techniques as detailed below were incorporated by creating a unique identification for each combination of patient and measurement technique, resulting in a total of 45 data sets for each of the two reading sessions (15 patients × 3 measurement techniques). The order of those 45 data sets was randomised for each reader to separate the three
analyses for each patient as well as to minimise a potential error due to learning effect within one reading session. Each reader was given a list only containing the measurement techniques attributed to the individual order from 1 to 45 for each reading session blinded to the patient identification.

Image analysis

Image analysis of all 90 data sets was performed by one vascular expert reader (3 years experience in vascular image post-processing, tutor in a Continuing Medical Education (CME)-certified vascular image post-processing course) and three non-expert readers (experience in <10 vascular image post-processing cases). Non-expert readers received 3 h of practical teaching by the expert reader. Furthermore, non-expert readers performed analyses of two training cases for each measurement technique under the supervision of the expert reader. For each reader, the 90 data sets prepared as described above were transferred to a commercially available image post-processing workstation (Aquarius, v.3.6.2.3, TeraRecon, Inc., San Mateo, CA, USA).

Two independent vascular expert radiologists established a standardised protocol for three different measurement techniques: manual measurements on axial source data, manual measurements using MPR and semi-automatic measurements using centerline analysis.

For all measurement techniques, four measurement positions in the aorta relevant to TEVAR were defined: P1, distal to left common carotid artery (or distal to bovine trunk in case of a bovine arch); P2, distal to left subclavian artery; P3, at the maximum diameter of the lesion; and P4, proximal to the coeliac trunk. Standard for window and levelling was 700/200 and was individually adjusted, if necessary. The target parameter was maximum diameter, at P1, P2 and P4 from inner wall to inner wall (including thrombus and excluding calcium), at P3 from outer wall to outer wall. In addition, the protocol for centerline analysis included length measurements from the left subclavian artery (P2) to the coeliac trunk (P4).

Statistical analysis

Data are given as mean ± SD. Measurements by the expert reader were defined as reference standard. Statistical analysis comprised four steps:

1. Analysis of original measurements. To exclude any systematic differences of expert measurements between the three techniques that might influence further statistical analysis, two-sided \( t \)-tests for paired samples were performed. It was tested whether expert manual adjustment of the sagittal, coronal and transverse planes to obtain an orientation perpendicular to the aorta (Fig. 1) and manual measurement. The protocol for centerline analysis was divided into two parts, preparation and measurements. The preparation of the centerline analysis was standardised following published terminology and recommendations: (1) placement of four seed points into the lumen centre (proximal ascending aorta, distal to the left subclavian artery, mid descending thoracic aorta and distal to the renal arteries); the software then computed the centerline along the vessel defined by the seed points; and (2) verification of the computed centerline and, if necessary, manual editing by optimisation of control points, adjustment of density thresholds or complete reset. Editing was deemed necessary if the centerline visibly deviated from the lumen centre or if the automatically reformatted images demonstrated any artefact. After editing, cross-sectional planes perpendicular to the aorta could be interactively viewed at any position. The software also automatically segmented the lumen and gave maximum, minimum and mean diameters for any position. For measurements, the appropriate reformatted perpendicular plane was identified (Fig. 2) and the automatically calculated maximum diameter was used for positions P1, P2 and P4, unless thrombus or calcium interfered with the automatic lumen segmentation. In this case and always at position P3, manual measurements were performed. The required time for all measurements was measured using a stopwatch. For centerline analysis, the time needed for preparation was measured separately. The centerline was subsequently used for measuring the length along the centerline from the left subclavian artery (P2) to the coeliac trunk (P4).
axial and centerline measurements significantly differed from expert MPR measurements, which are currently considered to be most accurate and reliable. Overall agreement between measurements by the non-expert readers and the reference standard was assessed for each measurement technique using Bland–Altman plots. The limits of agreement were calculated as mean ± 1.96 × SD.

(2) Reliability and inter-observer variability of measurements by non-expert readers. Reliability was defined as the absolute measurement deviation (AMD), determined as:

$$AMD = \left| \frac{M - R}{R} \right| \times 100\%,$$

with $M$ representing the measurement by the non-expert reader and $R$ representing the reference standard. Inter-observer variability was defined as the coefficient of variance (CV), calculated as:

$$CV = \frac{SD}{mean} \times 100\%,$$

with SD and mean representing the standard deviation and mean of the three measurements by non-expert readers. Analysis of variance was applied to test the impact of measurement technique, measurement position and slice thickness. In case of statistical significance, the differences for the subgroups of respective factors were assessed using a post hoc test with Fisher’s least significant difference (LSD) correction.

(3) Level of technical complexity. The most objective surrogate parameter for the level of technical complexity was considered to be the time needed by non-expert readers. Differences between the three measurement techniques were tested using analysis of variance and post hoc test with Fisher’s LSD correction.

(4) Training effect. Improvements of non-expert readers between the two reading sessions in AMD, CV and time difference to expert reader were assessed with Mann–Whitney U Test to determine a potential training effect.

A $p$-value of $\leq 0.05$ was considered to represent statistical significance. All analyses were performed with PASW Statistics Version 17.0 (SPSS Inc., Chicago, IL, USA).

**Results**

**Analysis of original measurements**

All included data sets were of good diagnostic quality. Image analysis according to the protocol was technically...
feasible for all data sets. Table 1 summarises diameter and length measurements of all four readers. Expert measurements showed good agreement between all three techniques. Expert measurements by axial and centerline technique did not significantly differ from expert measurements by the MPR technique ($p = 0.34$ and $0.31$). Fig. 3 gives examples of measurements for all techniques. For centerline analysis, manual measurements at P1, P2 and P4 were performed in 24% by the expert reader and in 18% by non-expert readers. Fig. 4 shows Bland–Altman plots and limits of agreement. Deviations from the reference standard for axial/MPR/centerline technique were $<1$ mm in 34%/36%/58% and $<3$ mm in 75%/79%/83%.

### Reliability and inter-observer variability

Tables 2 and 3 summarise AMD and CV calculations. Analysis of variance showed that both AMD and CV were significantly influenced by measurement technique ($p = 0.002$ and 0.009) and measurement position ($p < 0.001$ and $<0.001$). Slice thickness did not have a significant impact on AMD and CV ($p = 0.56$ and 0.49). Both AMD and CV were significantly lower for centerline analysis compared with axial technique ($p = 0.001$ and 0.042) and MPR ($p = 0.009$ and 0.003). Differences between axial technique and MPR were not significant ($p = 0.40$ and 0.33). Both AMD and CV were significantly higher at position P3 compared with positions P1 ($p < 0.001$ and $<0.001$), P2 ($p < 0.001$ and $<0.001$) and P4 ($p < 0.001$ and $= 0.002$). All other differences between positions were not statistically significant.

### Level of technical complexity

Table 4 summarises the time needed for analysis. Non-expert readers took significantly less time using the axial technique compared with MPR ($p < 0.001$) and centerline analysis ($p < 0.001$ for measurements only, $p < 0.001$ for total analysis). Analysis by centerline analysis was faster than by MPR ($p < 0.001$ for measurements only, $p = 0.55$ for total analysis).

### Training effect

AMD significantly improved in reading session 2 for MPR from 7.6% to 5.8% ($p = 0.007$). It did not show a significant change from the first to the second reading session for the axial technique (from 7.7% to 7.0%, $p = 0.86$) and for centerline analysis (from 4.8% to 4.6%, $p = 0.62$). CV improved in reading session 2 for MPR from 6.7% to 4.8%, but the improvement was not significant ($p = 0.17$). The variable did not show a relevant change for the axial technique (from 5.0% to 5.4%, $p = 0.08$) and for centerline analysis (from 4.1% to 3.7%, $p = 0.94$). The time difference between non-expert readers and the expert reader significantly decreased in reading session 2 for all techniques with a decrease from 2.2 to 0.4 min for the axial technique.
Discussion

This study demonstrated that reliability and inter-observer variability of preoperative measurements for TEVAR by non-expert readers were significantly improved by means of semiautomatic centerline analysis compared with manual assessment on axial slices and MPR. Use of semiautomatic centerline analysis resulted in better agreement with the reference standard, increasing the percentage of agreement of deviations <1 mm to 58% compared with 34% with the axial and 36% with the MPR technique. Measurements of maximum lesion diameter (P3) were significantly less reliable and significantly more variable than measurements defined by aortic branches for all measurement techniques, with semiautomatic centerline analysis still providing the best results. Slice thickness of 1 and 3 mm did not have a significant impact. Centerline analysis also enabled length measurements with excellent reliability and inter-observer variability. The time needed was highest for MPR followed by centerline analysis and the axial technique. In the second reading, there was a significant improvement for all techniques regarding time needed, but only for the MPR technique regarding measurement reliability.

Both centerline analysis and MPR aim at generation of cross sections perpendicular to the vessel course. Whereas MPR is a completely manual technique, centerline analysis provides computer-based assistance in two ways. First, perpendicular cross sections are automatically calculated based on the interactively created centerline. This eliminates the error associated with repeated manual adjustments of all three imaging planes as done in MPR. The present study shows that such repeated adjustments are time-consuming, reduce reliability and increase inter-observer variability among non-expert readers. As MPR is technically more complex, it offers room for improvement with growing experience, as demonstrated in this study. Second, centerline analysis provides automatic diameter calculations based on the preceding segmentation of the vessel lumen. The reader only has to check the accuracy of the automatic segmentation and to manually draw the diameter, if necessary. As the used software version did not provide the possibility for automated outer wall to outer wall measurements, diameters at position P3 had to be manually drawn in all cases. This resulted in reduced reliability and increased inter-observer variability at that position. In the future, this could be improved by automated outer wall to outer wall measurements.23

To our knowledge, this is the first study to assess different preoperative measurement techniques for TEVAR and the first study to investigate MPR. We demonstrated that measurements on axial slices in the thoracic aorta are similar to the abdominal aorta associated with considerable inter-observer variability.12,13 MPR has been suggested to improve accuracy compared with the axial technique in expert readers because diameters are measured perpendicular to the vessel course on MPR.4,15 The present study showed that, compared with measurements on axial slices, reliability among non-expert readers was increased using precision of measurements.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Reliability of measurements by the three techniques – absolute measurement deviations of non-expert readers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>Overall [%]</td>
</tr>
<tr>
<td>Axial</td>
<td>7.3 ± 7.7</td>
</tr>
<tr>
<td>MPR</td>
<td>6.7 ± 4.5</td>
</tr>
<tr>
<td>Centerline</td>
<td>4.7 ± 4.8</td>
</tr>
</tbody>
</table>

Summary of absolute measurement deviations (AMD) from reference standard given in [%] ± SD for each measurement technique and position (P1–P4, explanations see text).

(p < 0.001), from 2.8 to 0.4 min for MPR (p < 0.001) and from 3.0 to 0.7 min for centerline analysis (p < 0.001).
MPR at the expense of higher inter-observer variability and more time consumption. Computer-assisted measurements have previously been demonstrated to be less variable and less time-consuming compared with manual measurements for the abdominal aorta. The present study confirms this finding for the thoracic aorta. Irrespective of the measurement technique, aortic distension may have to be taken into account for endovascular stent-graft sizing, particularly in the thoracic aorta, where systolic and diastolic diameters can considerably differ.24 Length measurements, in principle, can be performed using all three presented techniques.4,25 However, length measurements by counting axial slices can only be accurate, if the aorta runs perpendicular to the axial slices. They are not plausible within the aortic arch and the descending thoracic aorta because counting axial slices will inevitably disregard the tortuous anatomy of the thoracic aorta.16,17 Furthermore, primary data are usually reconstructed before image archiving using an overlapping slice technique of 20–50% to reduce image noise. This will result in false length measurements, if the slice thickness is simply multiplied by the number of counted slices without considering the overlap. Length measurements by MPR involve assembling multiple straight measurements and thus also do not account for aortic tortuosity.4 Measurements constantly following the centerline are only possible using centerline analysis. For those reasons, the present study focussed on length measurements by centerline analysis. The excellent reliability and low inter-observer variability of length measurements by centerline analysis, as demonstrated in this study advocate its use in clinical practice. It has to be noted that the path of the endovascular stent-graft may not necessarily follow the centerline, particularly in aortic-arch pathologies.16 In such cases, an algorithm determining the length along the lesser curvature may be preferred.26

In everyday clinical practice, reader experience is an important factor that may influence measurement reliability. This study has two major clinical implications. First, centerline analysis should be used as measurement technique of choice by non-expert readers for preoperative assessment of TEVAR. Not only does centerline analysis provide high reliability and low inter-observer variability, but it also gives the possibility of easy and reliable length measurements along the computed centerline. Routine use of centerline analysis may improve stent-graft choice and, thus, reduce the incidence of postoperative complications.

The present study focussed on inter-observer variability for the following reason. Inter-observer variability is, like intra-observer variability, a measure of reproducibility, but inter-observer variability also includes errors related to different perceptions by different readers, which are of paramount importance in clinical routine. Finally, this study might be limited by the inclusion of scans with both 1- and 3-mm slice thickness. However, this was routine clinical scan protocol at the two CT scanners used for image acquisition. In addition, the present study showed that slice thickness did not have an impact on reliability and inter-observer variability.

In conclusion, this study demonstrates that semi-automatic centerline analysis provides the most reliable and least variable preoperative measurements for TEVAR among non-expert readers. MPR should be performed by expert readers due to significantly lower reliability and higher inter-observer variability among non-expert readers.

### Table 3

<table>
<thead>
<tr>
<th>Technique</th>
<th>Overall [%]</th>
<th>Compared to centerline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axial</td>
<td>5.2 ± 4.2</td>
<td>p = 0.042</td>
</tr>
<tr>
<td>MPR</td>
<td>5.8 ± 4.8</td>
<td>p = 0.003</td>
</tr>
<tr>
<td>Centerline</td>
<td>3.9 ± 5.4</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Summary of coefficients of variance (CV) of non-expert readers given in [%] ± SD for each measurement technique and position (P1–P4, explanations see text).

### Table 4

<table>
<thead>
<tr>
<th>Technique</th>
<th>Expert reader [min]</th>
<th>Non-expert readers [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axial</td>
<td>3.0 ± 0.5</td>
<td>4.3 ± 1.4</td>
</tr>
<tr>
<td>MPR</td>
<td>5.5 ± 0.9</td>
<td>7.2 ± 2.1</td>
</tr>
<tr>
<td>Centerline</td>
<td>5.1 ± 1.6</td>
<td>7.0 ± 2.3</td>
</tr>
<tr>
<td>Preparation only</td>
<td>1.4 ± 0.5</td>
<td>1.9 ± 0.7</td>
</tr>
<tr>
<td>Measurements only</td>
<td>3.7 ± 1.4</td>
<td>5.0 ± 1.9</td>
</tr>
</tbody>
</table>
Use of centerline analysis in everyday clinical practice may improve stent-graft choice and, consequently, the incidence of postoperative complications.

Conflict of Interest/Funding
None.

Acknowledgements
Fabian Rengier received a grant from the German Research Foundation within the 'Research training group 1126: Intelligent Surgery — Development of computer-based methods for the future workplace in surgery’. We thank Lisa Zipp and Maximilian Fuchs for participating in the data evaluation.

References