Inter-observer Variability in Sizing Fenestrated and/or Branched Aortic Stent-grafts

H. Banno [a,*], H. Kobeiter [b], J. Brossier [a], J. Marzelle [a], E. Presles [c], J.-P. Becquemin [a]

[a] Department of Vascular Surgery, Henri Mondor Hospital, University Paris XII, Creteil, France
[b] Department of Radiology, Henri Mondor Hospital, University Paris XII, Creteil, France
[c] Department of Clinical Research, Saint-Étienne University Hospital, St-Étienne, France

WHAT THIS PAPER ADDS

The current study is the first large-scale analysis, focused on inter-observer variability in sizing fenestrated and/or branched aortic stent-grafts. The agreements between core laboratory and each rater were all moderate to perfect; however, there were some significant discrepancies, which may affect clinical results. These discrepancies should be taken into account in sizing fenestrated and/or branched stent-grafts.

Background: Several studies have examined inter-observer variability in measurements for standard EVAR, but little is known about measurements for complex aortic aneurysm.

Methods: Two independent observers reviewed all preoperative CT scans of 268 patients in a French trial of fenestrated and/or branched aortic stent-grafts (f/b-EVAR). Those data were compared with those obtained (1) by investigators (extent of aneurysm, target vessel stenosis, and aortic diameters), and (2) from manufacturers (proximal landing zone, device diameter, and target vessel position). We assessed the reproducibility using kappa statistics for qualitative data and both Bland—Altman plot and Passing—Bablok regression analysis for quantitative data.

Results: Reproducibility was moderate to almost perfect for all factors. However, a few critical discrepancies were found, such as target vessel clock position ($\geq$45 minutes) and location ($\geq$5 mm), level of proximal landing zone, and diameters of the endograft.

Conclusions: This is the first large-scale analysis focused on inter-observer variability in sizing for f/b-EVAR. The measurement data showed good agreement, but there were some critical discrepancies between observers that may affect clinical results.

Article history: Received 10 July 2013, Accepted 7 October 2013, Available online 23 October 2013

Keywords: Aortic aneurysm, EVAR, Fenestrated endovascular aneurysm repair, Inter-observer variability, Sizing

Complex endovascular repair, such as fenestrated and/or branched endovascular aortic repair (f/b-EVAR), is a recent development.[1–3] Although commercially available standardized devices have been developed, at present custom-made devices are used in the mainstream in this area.[4–6] Whereas standardized devices are designed to be suitable for a certain average anatomy, custom-made devices require accurate preoperative sizing of stent-grafts for technical success. The design of a custom-made device is based on an individual CT scan provided by a surgeon. Device planning requires experience in imaging and 3D reconstruction using a workstation to make all the necessary measurements. In the majority of cases, this sizing is performed by specialists in a centralized planning facility of the manufacturer.

Inter-observer variability in various exams is well known. Some authors have reported inter-observer variability in measurements of abdominal aortic aneurysm (AAA).[7,8] However, at present, little is known about any discrepancy of sizing of complex aortic stent-grafts between different specialists and between clinicians and manufacturers. This study investigates the variability between experienced endovascular surgeons and investigator or manufacture measurements in measuring and sizing endovascular aneurysm repair using fenestrated and/or branched stent-grafts.

METHODS

WINDOWS study CT scans

WINDOWS Study is a multicenter, prospective single-arm trial of f/b-EVAR for complex aortic aneurysms — abdominal (juxta-, para-, and suprarenal AAA) or thoracoabdominal (TAAA) — in centers selected according to

* Corresponding author. H. Banno, Departments of Vascular Surgery, Henri Mondor Hospital, University Paris XII, Creteil, France.
E-mail address: hiroshib@gmail.com (H. Banno).

1078-5884/$ — see front matter © 2013 European Society for Vascular Surgery. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ejvs.2013.10.008

their expertise in this technique and their compliance with the recommendations of the French Health Authority (HAS: Haute Autorité de Santé). All patients had preoperative CT scans and patient inclusion was validated by both the inclusion criteria committee of WINDOWS study and the planning center of manufacturer. Between September 2009 and October 2012, 268 patients were included in the trial (the study is registered # NCT01168037 at clinicaltrials.gov (http://www.clinicaltrials.gov/ct2/show/NCT01168037)).

In this study, all the preoperative CT scans were collected and reviewed by the core laboratory. Planning center data of the manufacturer were collected, as well as data provided by investigators. The quality of the retrieved scans varied widely in terms of slice thickness (1 mm to 5 mm) and scanning interval after contrast injection, and thus not all scans were optimal for sizing fenestrated and/or branched endograft.

Image analysis

Two independent observers performed image analysis as the core laboratory. A three-dimensional imaging workstation (TeraRecon Inc., Santa Rosa, CA, USA.) was used to generate multiple three-dimensional reconstructions of volumetric data sets from the preoperative CT scans. Both observers were well-trained and experienced vascular surgeons. The third observer, an experienced interventional radiologist, provided the final decision as a core laboratory in case of discrepancy in categorization between the two observers. As for the quantitative data, mean values of the two observers were determined as core laboratory data.

Extent of aneurysm was classified according to reporting standards for thoracic endovascular aortic repair (TEVAR) and ACC/AHA guidelines.[9,10] As for paravisceral aneurysm, juxtarenal aneurysms arise distal to the renal arteries but in very close proximity to them; pararenal aneurysms involve the origin of one or both renal arteries; suprarenal aneurysms encompass the visceral aortic segment containing the superior mesenteric and celiac arteries.

Eventual stenosis ($>$70%) of visceral branches (celiac axis [CA], superior mesenteric artery [SMA], right renal artery [RRA], left renal artery [LRA]) was identified. Stenosis determination was made by measuring the ratio between the diameter of the narrowest segment of the imaged artery (a) and the diameter of a normal segment of the artery proximal to the stenosis or distal to poststenotic dilation (b) (Percentage of stenosis $=$ (b $-$ a)/b $\times$ 100).

A semi-automated centerline was generated using the above-mentioned workstation. The centerline was assessed with multiplanar reconstruction views perpendicular to the centerline of flow, and then manually edited if necessary. Aortic diameters at each level of visceral branches (CA, SMA, RRA, and LRA), thoracic and infrarenal aortic diameter were measured in perpendicular planes to the centerline.

Visceral artery orientation was measured relative to a line extending anteriorly from the centerline of the aorta. Clockwise deviation was assigned a positive value, and counterclockwise deviation a negative value. The average of angles estimated by two observers was defined as the angle of core laboratory. And then, all degrees were converted to clock positions for analysis considering 0° as 12 o'clock because some data about target vessel orientation obtained from manufacturer were described only as clock positions.

For measuring longitudinal vessel separation, a stretch view was used. The distance between the center of each target vessel ostium and the low margin of CA ostium was measured. In the case when information of CA could not be obtained, the low margin of SMA ostium was substituted as a reference point.

The proximal aorta was considered to be suitable as a landing zone when the length of healthy aorta was $\geq$15 mm. Aneurysms were sub-divided into zones according to where it was thought an adequate proximal seal could be achieved in relation to the visceral arteries. Zone 0 was a seal below the lowest preserved renal artery, Zone 1 is between renal arteries at different levels, Zone 2 was above the renal arteries but below the SMA, Zone 3 was above the SMA but below the CA and Zone 4 was above the CA (Dr. K. Ivancev, personal communication, June 2013).

The proximal device diameter was determined according to the aortic diameter in the proximal seal zone and in agreement with the instructions for use of the manufacturer.

Comparing data

Data about extent of aneurysm, stenosis of visceral branches, and aortic diameter, were obtained from each center. They were estimated or measured by their own way in daily practice. Orientation of visceral arteries, distance from low margin of CA (or SMA), and proximal device diameter were obtained from the manufacturer. The proximal seal zone that the manufacturer proposed was obtained from the planning sheet of the manufacturer. (A circumferential seal was expected at the level of fenestration but scallop, which means that the proximal landing zone was considered distal to the scallop if the device incorporated a scallop.)

Data analysis

All statistical analyses were performed using SAS version 9.3 (SAS Institute, Cary, NC, USA) or R statistical software, version 3.0.0 (A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). Quantitative and qualitative variables were analyzed separately by several methods. For quantitative variables, agreement between the core laboratory and raters was assessed by plotting the difference between each reading and the reference with the limits of agreement ($\pm$two standard deviations around the mean difference) as described by Bland and Altman.[11] Quantitative variables were also analyzed by Passing–Bablok regression.[12] For qualitative variables, reproducibility was assessed using the weighted kappa statistics (quadratic weighting was employed). Applying generally accepted definitions, kappa values $\leq$0 indicate no agreement, 0 to 0.2 slight agreement,