

**Are your MRI contrast agents cost-effective?**

Learn more about generic Gadolinium-Based Contrast Agents.



**AJNR**

**Diagnostic Accuracy of 4 Commercially Available Semiautomatic Packages for Carotid Artery Stenosis Measurement on CTA**

J. Borst, H.A. Marquering, M. Kappelhof, T. Zadi, A.C. van Dijk, P.J. Nederkoorn, R. van den Berg, A. van der Lugt and C.B.L.M. Majoie

This information is current as of April 19, 2024.

*AJNR Am J Neuroradiol* 2015, 36 (10) 1978-1987

doi: <https://doi.org/10.3174/ajnr.A4400>

<http://www.ajnr.org/content/36/10/1978>

# Diagnostic Accuracy of 4 Commercially Available Semiautomatic Packages for Carotid Artery Stenosis Measurement on CTA

J. Borst, H.A. Marquering, M. Kappelhof, T. Zadi, A.C. van Dijk, P.J. Nederkoorn, R. van den Berg, A. van der Lugt, and C.B.L.M. Majoie



## ABSTRACT

**BACKGROUND AND PURPOSE:** Semiautomatic measurement of ICA stenosis potentially increases observer reproducibility. In this study, we assessed the diagnostic accuracy and interobserver reproducibility of a commercially available semiautomatic ICA stenosis measurement on CTA and estimated the agreement among different software packages.

**MATERIALS AND METHODS:** We analyzed 141 arteries from 90 patients with TIA or ischemic stroke. Manual stenosis measurements were performed by 2 neuroradiologists. Semiautomatic measurements by using 4 methods (3mensio and comparable software from Philips, TeraRecon, and Siemens) were performed by 2 observers. Diagnostic accuracy was estimated by comparing semiautomatic with manual measurements. Interobserver reproducibility and agreement between different packages was assessed by calculation of the intraclass correlation coefficient and Bland-Altman 95% limits of agreement. False-negative classifications were retrospectively inspected by a neuroradiologist.

**RESULTS:** There was no significant difference in the diagnostic performance of the 4 semiautomatic methods. The sensitivity for detecting  $\geq 50\%$  and  $\geq 70\%$  degree of stenosis was between 76% and 82% and 46% and 62%, respectively. Specificity and overall diagnostic accuracy were between 92% and 97% and 85% and 90%, respectively. The interobserver intraclass correlation coefficient was between 0.83 and 0.96 for semiautomatic measurements and 0.81 for manual measurement. The limits of agreement between each pair of semiautomatic packages ranged from  $-18\%$ – $24\%$  to  $-33\%$ – $31\%$ . False-negative classifications were caused by ulcerative plaques and observer variation in stenosis and reference measurements.

**CONCLUSIONS:** Semiautomatic methods have a low-to-good sensitivity and a good specificity and overall diagnostic accuracy. The high interobserver reproducibility makes semiautomatic stenosis measurement valuable for clinical practice, but semiautomatic measurements should be checked by an experienced radiologist.

**ABBREVIATION:** ICC = intraclass correlation coefficient

Carotid endarterectomy in neurologically symptomatic patients with a 70%–99% stenosis results in a 16% decrease in the absolute risk for an ipsilateral stroke in 5 years. However, endarterectomy is only marginally beneficial for patients with a

50%–69% stenosis and has no positive effect in patients with a  $<50\%$  stenosis.<sup>1</sup> Therefore, the degree of carotid stenosis is crucial in clinical decision-making, and precise and accurate measurement of the degree of stenosis is mandatory. The stenosis measurements on which these thresholds are based were determined by using conventional angiography, which is considered as the original criterion standard.<sup>2</sup> Due to neurologic complications related to DSA<sup>3</sup> and a good diagnostic accuracy of noninvasive tests, carotid stenosis measurement on CTA or MRA has become the standard in clinical practice.<sup>4,5</sup> However, manual measurement of the degree of stenosis on CTA according to the NASCET method is prone to low interobserver reproducibility and requires experience.<sup>6,7</sup> Semiautomatic methods increase the interobserver reproducibility and accelerate the measurement.<sup>8,9</sup> Furthermore, semiautomatic methods require less observer experience compared with manual measurement.<sup>10</sup> Multiple semiautomatic packages are currently available and used in clinical practice. Be-

Received December 17, 2014; accepted after revision February 26, 2015.

From the Departments of Radiology (J.B., H.A.M., M.K., R.v.d.B., C.B.L.M.M.), Biomedical Engineering and Physics (H.A.M.), and Neurology (P.J.N.), Academic Medical Center, Amsterdam, the Netherlands; and Department of Radiology (T.Z., A.C.v.D., A.v.d.L.), Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands.

This work was supported by the Information Technology for European Advancement (ITEA)2 project, label ITEA 10004: Medical Distributed Utilization of Services & Applications.

Please address correspondence to Jordi Borst, MSc. MD, Department of Radiology, GI-229, Academic Medical Center Amsterdam, Meibergdreef 9, 1105AZ Amsterdam, the Netherlands; e-mail: j.borst@amc.uva.nl



Indicates article with supplemental on-line tables.



Indicates article with supplemental on-line photo.

<http://dx.doi.org/10.3174/ajnr.A4400>

cause different vendors may use different algorithms,<sup>11</sup> the reliability of measurements with different software packages is unclear. To become a valuable clinical tool, the diagnostic accuracy must be further investigated. The goal of this study was to assess the agreement and diagnostic accuracy of 4 commercially available software packages for semiautomatic stenosis measurement compared with manual measurement on CTA and to estimate the interobserver reproducibility and the agreement among different semiautomatic packages.

## MATERIALS AND METHODS

### Patient Selection

Patients with a recent TIA or stroke suspected of having ICA stenosis were evaluated by duplex sonography. According to local guidelines, when the stenosis on duplex sonography was  $\geq 30\%$  for a man and  $\geq 50\%$  for a woman, CTA was performed to estimate the degree of stenosis more precisely. All consecutive patients ( $n = 110$ ) who underwent a 64-section CTA with a 0.9-mm section thickness for carotid stenosis evaluation between April 2006 and December 2008 were retrospectively included in this analysis. This complete population was previously investigated to assess the performance of semiautomatic measurement of ICA stenosis on CTA by using Vitrea 2 version 4.1.2.0 (Vital Images, Plymouth, Minnesota).<sup>8</sup> In the current study, we report the diagnostic accuracy and reproducibility of semiautomatic carotid stenosis measurement on CTA by using 4 other commercially available software packages and estimate the agreement among different software packages. Furthermore, this complete population was previously used to investigate the relation of calcium volume with carotid artery disease,<sup>12</sup> and it was also used to investigate the prevalence of intracranial carotid artery disease and quantify the intracranial stenosis,<sup>13,14</sup> and to investigate the relation between intracranial carotid artery stenosis and poor outcome.<sup>15</sup>

Patients with a previous carotid intervention ( $n = 16$ ) and those with CTA of insufficient quality ( $n = 4$ ) were excluded; 90 patients remained for further analysis. The mean age was 66.8 years (range, 35–89 years), and 54 were men. Forty patients (44%) had ischemic stroke as a final diagnosis; 32 patients (36%), a transient ischemic attack; and 14 patients (16%), amaurosis fugax. Three patients (3%) were asymptomatic, and 1 patient (1%) had an ocular ischemic syndrome.

Because CTA was performed in the clinical setting, informed consent was waived by the local medical ethics committee.

### CTA Protocol

CTA was performed as previously described<sup>8</sup> with a 64-section scanner (Brilliance 64; Philips Healthcare, Best, the Netherlands). Eighty milliliters of contrast (iodixanol, Visipaque 320; GE Healthcare, Piscataway, New Jersey) was infused at 4 mL/s. Acquisition and reconstruction parameters were as follows: 120-kV tube voltage, 265 mAs, pitch of 0.765, and reconstructed section thickness of 0.9 mm with an increment of 0.45 mm. The scan ranged from the aortic arch up to 3 cm above the sella turcica. The in-plane grid was  $512 \times 512$  pixels, with an FOV ranging from  $128 \times 128$  mm<sup>2</sup> to  $217 \times 217$  mm<sup>2</sup>, with an average of  $155 \times 155$  mm<sup>2</sup>.

### Stenosis Measurement

For both the manual and semiautomatic measurements, the observers were blinded to patient information and each other's findings. The degree of stenosis was defined according to the NASCET criteria<sup>2</sup> by using the minimal diameter at the stenosis and the maximum reference diameter at a healthy part of the artery well beyond ( $>30$  mm) the stenosis.<sup>8</sup> Because the cross-section of an artery is not round, there is no true diameter. Therefore, we defined the "minimal diameter" as the minimal cross-sectional distance of the artery from wall to wall and the "maximum diameter" as the maximum cross-sectional distance of the artery from wall to wall. The minimal diameter of the stenosis was determined by the observers within 3 cm proximal and distal to the bifurcation.<sup>16</sup> Arteries with near-occlusion (collapsed or small distal artery) were identified according to the criteria described by Bartlett et al.<sup>17</sup> For both the manual and semiautomatic measurements, occlusion of the arteries was reported. For all measurements, the processing time was recorded.

### Manual Stenosis Measurements

Manual measurements were performed on CTA by 2 neuroradiologists both with  $>10$  years of experience according to the method described by Bartlett et al<sup>17</sup> by using a workstation with MPR functionality (Impax, Version 5.2; Agfa-Gevaert, Mortsel, Belgium). Measurements were performed on a plane perpendicular to the centerline of the artery. The first observer measured all arteries, which were used as the reference, and a subset of 50 arteries a second time with a delay of 2 months. The second observer measured a subset of 48 arteries.

### Semiautomatic Stenosis Measurements

Semiautomatic stenosis grading was performed with software packages from Pie Medical Imaging (3mensio Vascular 6.1; Pie Medical Imaging Maastricht, the Netherlands), Philips (Extended Brilliance Workspace, Version 4.1 Advanced Vessel Analysis), TeraRecon (Vessel Analysis 4.4.6.85; TeraRecon, San Mateo, California), and Siemens (syngo INspace4D Advanced Vessel Analysis 2009–2013; Siemens, Erlangen, Germany). One trained observer (2 years of experience) performed stenosis measurement by using all software packages with  $>2$  months between measurements with different packages. A second trained observer (6 months of experience) performed the measurements by using Philips and 3mensio software, and a third trained observer (6 months of experience) performed the measurements by using Siemens and TeraRecon software, both with  $>2$  months between measurements with different software packages. To prevent recall of measurements performed in previous studies on this population, we selected observers who were not involved in the previous studies.<sup>8,12–15</sup>

Using the software from Philips, TeraRecon, and Siemens, we placed  $\geq 2$  seed points on the axial images: The first seed point was placed in the ICA close to the base of the skull, and the last seed point, in the common carotid artery below the bifurcation ( $>5$  cm). Subsequently, software packages automatically segmented the ICA and determined the centerline of the ICA. Minimal and maximal lumen diameters of the arteries were automatically calculated and displayed together with

curved planar reformations of the artery. By dragging a slider along the curved planar reformation of the artery, observers were able to select the minimal stenosis diameter. For the 3mensio software, the ICA of interest was automatically segmented after placement of a single seed point. Subsequently, seed points were placed in the ICA, bifurcation, external carotid artery, and common carotid artery on a 3D representation, and the centerline was automatically determined. The 3mensio software fitted an ellipse on the segmented cross-sectional lumen area and presented the minimal and maximum diameters of the ellipse as the lumen diameters and displayed them together with curved planar reformations of the artery. The observer selected the region of the ICA containing the stenosis, and the software automatically determined the smallest diameter of the stenosis. For all software packages, the reference location was selected by dragging a slider on the curved planar reformation along the distal ICA well beyond the site of stenosis. The reference location was selected at a vertically running part with the largest diameter and the least variation in diameter. At the selected reference location, the minimal and maximal diameters were recorded. For all software packages, erroneous or incomplete segmentations and erroneous centerlines were manually corrected.

To evaluate potential improvements of the interobserver reproducibility, we performed an additional measurement with a second standardized reference location exactly 30 mm above the minimal stenosis diameter for a single software package (Siemens) and calculated the degree of stenosis.

### Statistical Analysis

**Diagnostic Accuracy.** To determine the diagnostic accuracy of semiautomatic stenosis measurement, we used the manual stenosis measurements by the first observer as a reference. The agreement of the semiautomatic stenosis measurements with the manual reference was assessed by scatterplots, Bland-Altman analysis with 95% limits of agreement, and the calculation of the intraclass correlation coefficient (ICC) (agreement, 2-way-mixed, single measure). Diagnostic accuracy was determined for diagnoses of  $\geq 50\%$  and  $\geq 70\%$  stenoses. Sensitivity, specificity, positive predictive value, negative predictive value, and overall diagnostic accuracy were calculated. The extended McNemar test was used to compare the sensitivity, specificity, and overall diagnostic accuracy among the software packages. *P* values  $< .05$  were considered statistically significant. Statistical analyses were performed by using SPSS, Version 21 (IBM, Armonk, New York).

**Inter- and Intraobserver Reproducibility.** Inter- and intraobserver reproducibility of the manual measurements and interobserver reproducibility of the semiautomatic measurements were assessed by scatterplots, Bland-Altman analysis, and the calculation of the ICC. A paired *t* test was used to determine whether the interobserver bias for semiautomatic measurements was statistically different from manual measurements. The Fisher Z-test was used to determine whether the interobserver ICC for semiautomatic measurements was statistically significantly different from manual measurements. The agreement of observers classifying a

**Table 1: Average processing time**

	Average Processing Time $\pm$ SD (seconds)
Manual measurements	138 $\pm$ 31
3mensio	86 $\pm$ 42
Philips	115 $\pm$ 77
TeraRecon	84 $\pm$ 64
Siemens	89 $\pm$ 86

stenosis equal to or higher than a cutoff of 50% and 70% was assessed by using  $\kappa$  statistics.

### Agreement between Different Semiautomatic Software Packages.

The agreement between measurements with different semiautomatic software packages was assessed by Bland-Altman analysis and calculating the ICC. Instead of choosing a fixed observer per software package, we randomly selected 1 of the 2 observers for each measurement to avoid observer dependence. Thus, we aimed to simulate a clinical setting in which multiple users may use the software package.

### Retrospective Error Analysis

Semiautomatic measurements classified as false-negative were retrospectively investigated by a neuroradiologist (10 years of experience) and trained observer (2 years of experience) to inspect whether the measurement was correctly performed by the observers and no erroneous centerlines or erroneous lumen segmentations were present. A measurement was classified as false-negative if the degree of stenosis was above the cutoff point (50% or 70%) according to manual measurement but below the cutoff for the semiautomatic measurement.

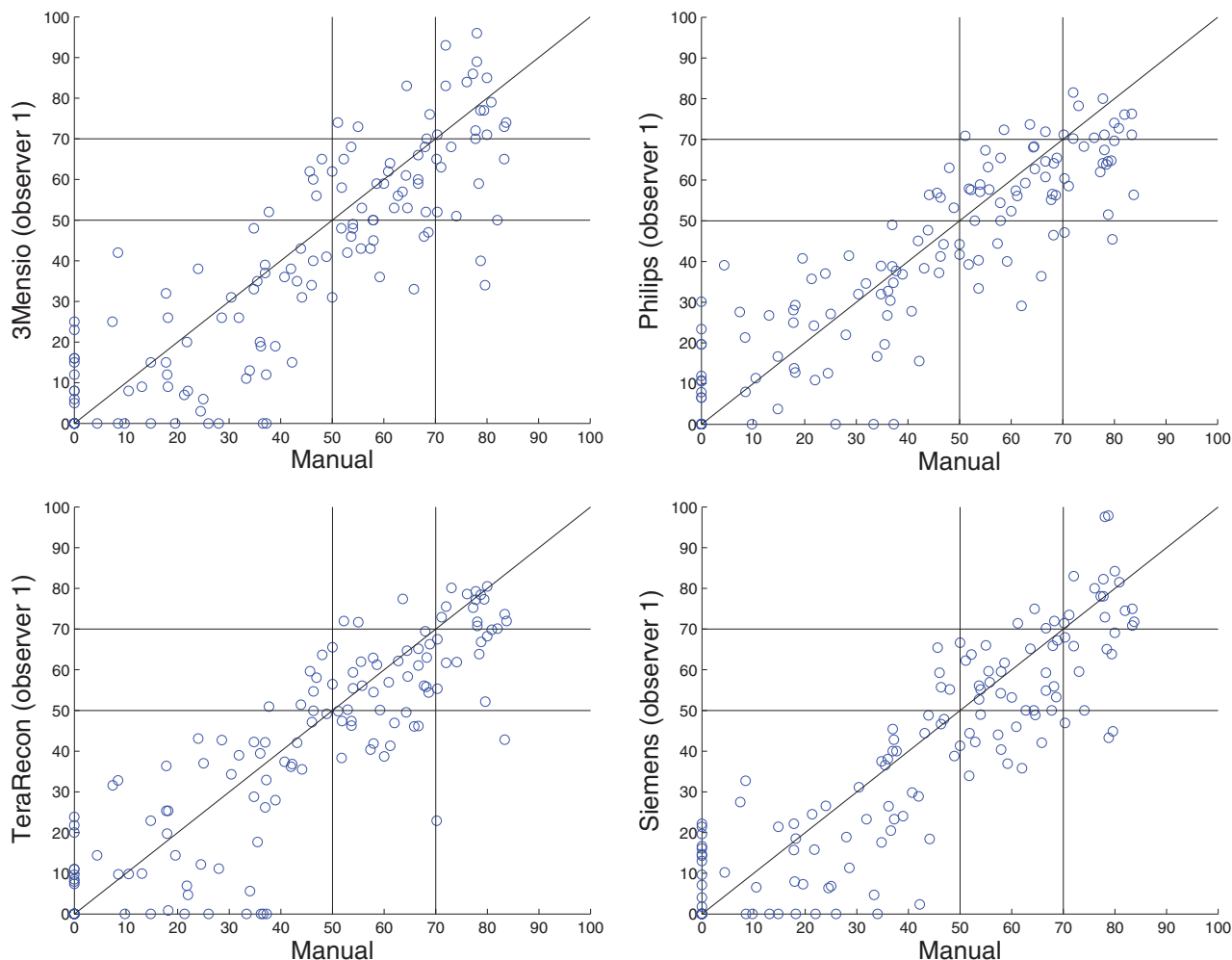
## RESULTS

Ninety patients (180 arteries) were included in this study. Thirty-nine arteries were excluded because of near-occlusion ( $n = 20$ ), occlusion ( $n = 13$ ), dental artifacts at the bifurcation ( $n = 3$ ), dissection ( $n = 1$ ), and fibromuscular dysplasia ( $n = 1$ ), or the bifurcation was not captured on the scan ( $n = 1$ ). After exclusion, we ended up with 141 (180–39) arteries suitable for further analysis. A subset of 38 arteries that were manually measured a second time by the first observer and 37 arteries that were manually measured by the second observer were suitable for further analysis. According to the manual stenosis measurements, 47 arteries had a minimal stenosis (0%–29%); 29, a mild stenosis (30%–49%); 39, a moderate stenosis (50%–69%); and 26, severe stenosis (70%–99%).

As Table 1 shows, the average processing time of all semiautomatic measurements was faster than that for manual measurements. See On-line Fig 1 for examples of semiautomatic ICA stenosis measurement.

### Diagnostic Accuracy

The agreement of semiautomatic measurements with manual measurements is illustrated in Fig 1 by scatterplots. The ICC and limits of agreement are shown in Table 2. All software packages showed a high correlation, with ICCs between 0.86 and 0.88. The mean paired difference between manual and semiautomatic measurements was small, ranging from  $2.1\% \pm 13\%$  to  $3.8\% \pm 14\%$  (Fig 2). However,



**FIG 1.** Scatterplots of the manual-versus-semiautomatic assessment of the degree of stenosis (percentage).

**Table 2: Agreement manual vs semiautomatic stenosis measurement**

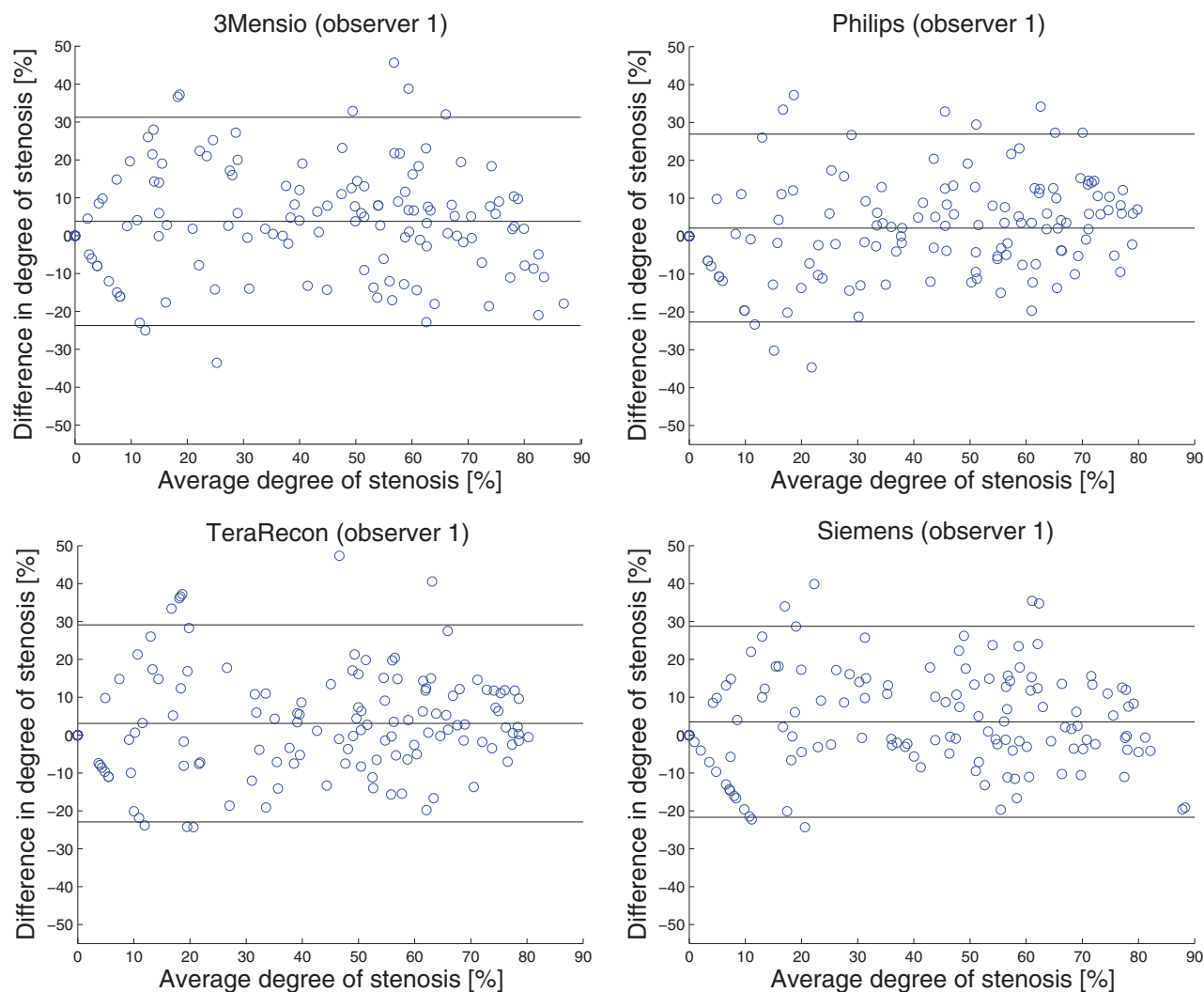
	Average Difference Degree of Stenosis $\pm$ SD (%) (Manual, Semiautomatic)	Bland-Altman 95% Limits of Agreement (%)	ICC for Degree of Stenosis (95% CI)
3mensio (observer 1)	$3.8 \pm 14$ ( $P = .002$ )	−24–31	0.86 (0.80–0.90)
Philips (observer 1)	$2.1 \pm 13$ ( $P = .049$ )	−23–27	0.88 (0.83–0.91)
TeraRecon (observer 1)	$3.1 \pm 13$ ( $P = .007$ )	−23–29	0.87 (0.82–0.90)
Siemens (observer 1)	$3.5 \pm 13$ ( $P = .002$ )	−22–29	0.88 (0.83–0.91)

the Bland-Altman limits of agreement were wide, ranging from −23%–27% to −24%–31% (Fig 2). The diagnostic performance is presented in Table 3. The semiautomatic measurements have a low sensitivity for detecting a  $\geq 70\%$  stenosis, with sensitivity values between 46% and 62%. The specificity and overall diagnostic accuracy of detecting  $\geq 70\%$  degree stenosis were good for semiautomatic measurements, ranging between 96%–97% and 87%–90%, respectively. The semiautomatic measurements showed a moderate-to-good sensitivity for detecting  $\geq 50\%$  stenosis with values between 68% and 82%. The specificity and overall diagnostic accuracy for detecting  $\geq 50\%$  stenosis were good, ranging between 93%–95% and 85%–88%, respectively. No statistically significant differences in the diagnostic performance among software packages were found. All occluded arteries were detected by the observers regardless of the semiautomatic method used.

### Inter- and Intraobserver Reproducibility

Observer reproducibility is illustrated in Figs 3–5, and the results can be found in Tables 4 and 5. The Bland-Altman plots showed a small inter- and intraobserver reproducibility bias with wide limits of agreement for manual stenosis measurements (Fig 3). The manual measurements have a reasonable-to-good inter- and intraobserver reliability, with an ICC of 0.81 and 0.88, respectively. The Bland-Altman plots show that interobserver reproducibility bias was smallest for 3mensio and Philips (Fig 5). The semiautomatic measurements have a reasonable-to-excellent interobserver reproducibility with ICCs between 0.83 and 0.96. For 3mensio and Philips, the interobserver reproducibility was significantly better than the interobserver reproducibility of the manual measurements. With the Siemens software with a fixed reference location 3 cm above the minimal stenosis diameter, the average difference in degree of steno-





**FIG 2.** Bland-Altman plots of the degree of stenosis determined by manual and semiautomatic assessment. The black lines represent the mean paired difference and 95% limits of agreement. The characteristic V-shape in the Bland-Altman plot is caused by 1 of the 2 measurements being zero with the other measurement being nonzero. These measurements happened particularly when the degree of stenosis was small ( $<30\%$ ).

**Table 3: Diagnostic performance of semiautomatic measurement in detecting a stenosis degree of  $\geq 70\%$  and  $\geq 50\%$**

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
<b>70% Cutoff</b>					
3mensio (observer 1)	62 (43–78)	96 (91–98)	76 (53–92)	92 (85–96)	89 (82–93)
Philips (observer 1)	46 (29–65)	96 (91–99)	75 (48–93)	88 (82–94)	87 (80–92)
TeraRecon (observer 1)	58 (37–77)	97 (93–99)	83 (59–96)	91 (85–95)	90 (84–94)
Siemens (observer 1)	62 (40–81)	97 (91–99)	80 (56–94)	92 (85–96)	90 (84–94)
<b>50% Cutoff</b>					
3mensio (observer 1)	77 (65–86)	93 (86–97)	91 (80–97)	83 (73–90)	86 (79–91)
Philips (observer 1)	82 (70–89)	93 (86–97)	91 (81–97)	86 (76–92)	88 (81–93)
TeraRecon (observer 1)	76 (65–86)	92 (84–97)	89 (78–96)	82 (73–90)	85 (78–90)
Siemens (observer 1)	77 (65–86)	95 (87–99)	93 (82–98)	83 (73–90)	87 (80–91)

**Note:**—PPV indicates positive predictive value; NPV, negative predictive value.

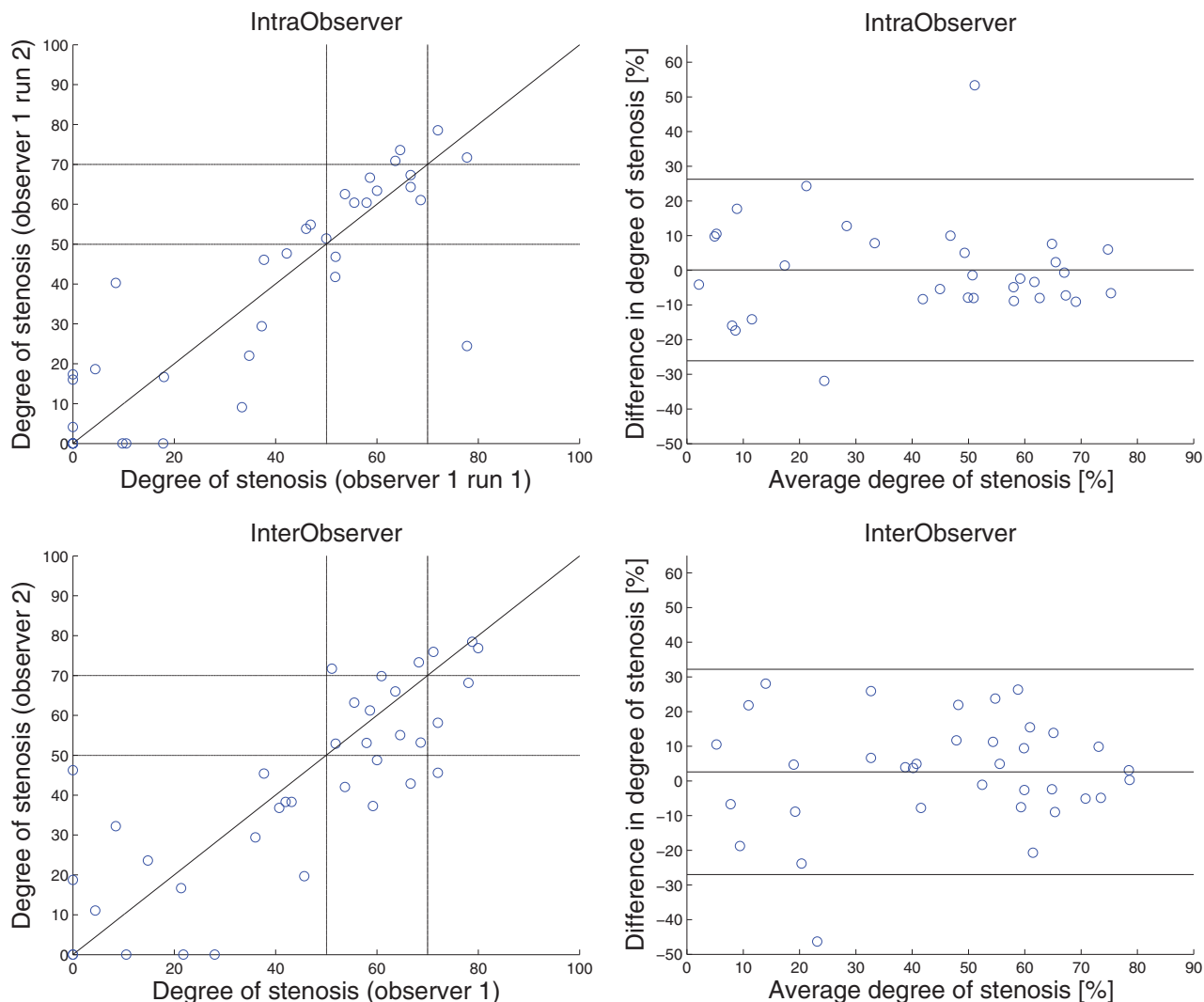
sis was  $3.5\% \pm 15\%$  compared with  $6.5\% \pm 12\%$  for the standard reference location ( $P < .001$ ) and the interobserver reproducibility was slightly lower, with an ICC of 0.84 compared with 0.86 with a non-statistically significant difference ( $P = .55$ ).

For detecting a stenosis of  $\geq 50\%$ , the  $\kappa$  statistics for the interobserver agreement were good for manual measurement and, depending on the software package, fair to excellent for the semiautomatic measurements (Tables 4 and 5). For detecting a stenosis of  $\geq 70\%$ ,

the  $\kappa$  statistics for interobserver agreement were fair for the manual measurement and, depending on the software package, poor to excellent for the semiautomatic measurement packages.

#### **Agreement among Semiautomatic Measurements**

The agreement among measurements with different semiautomatic software packages can be found in the On-line Table. The correlation of measurements with different semiautomatic pack-



**FIG 3.** Scatterplot (upper left corner) and Bland-Altman plot (upper right corner) of the repeated manual stenosis measurement (percentage) (intraobserver). Scatterplot (lower left corner) and Bland-Altman plot (lower right corner) of the manual assessment of the degree of stenosis (percentage) measured by observer 1 and observer 2 (interobserver). The black lines in the right figures represent the mean paired difference and 95% limits of agreement.

ages is high, with ICCs ranging from 0.92 to 0.98. The mean paired differences between semiautomatic packages range from 0.49% to 5.7%, and the Bland-Altman limits of agreement are wide, ranging from  $-17\%$ – $18\%$  to  $-33\%$ – $31\%$ .

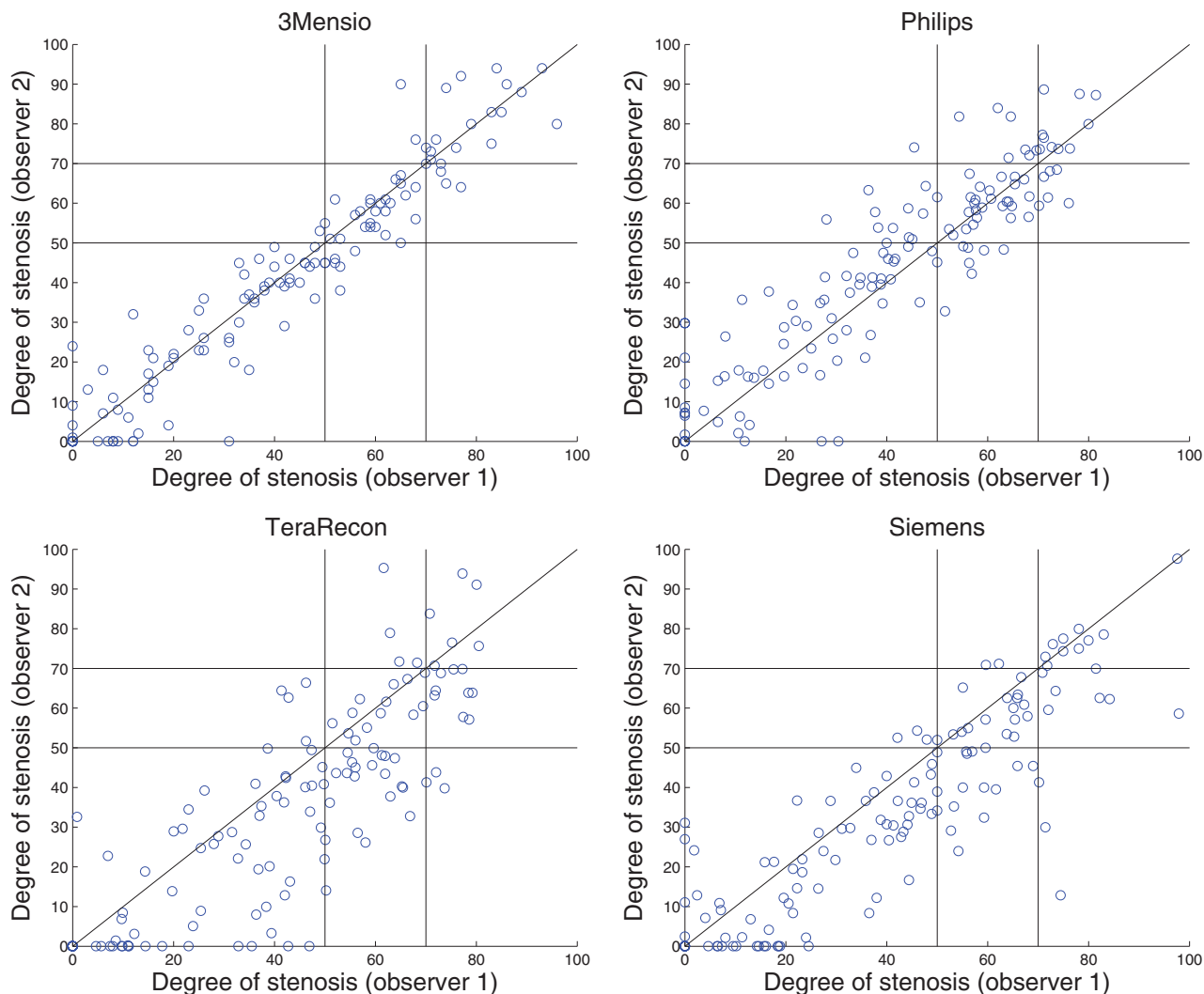
### Retrospective Error Analysis

Most measurements classified as false-negative were because the semiautomatic method measured a larger stenosis diameter and/or a smaller reference diameter compared with manual measurements by the neuroradiologist (78% [36/46] for a stenosis of  $\geq 70\%$  and 89% [51/57] for a stenosis of  $\geq 50\%$ ). There were no apparent errors in the centerline, and only 4.3% (2/46) of the false-negatives for a stenosis of  $\geq 70\%$  and 5.3% (3/57) for a stenosis of  $\geq 50\%$  were caused by erroneous lumen segmentation due to calcium. An ulcerative plaque hampered semiautomatic measurements in 17.4% (8/46) for a stenosis of  $\geq 70\%$  and 5.3% (3/57) for a stenosis of  $\geq 50\%$  and resulted in severe overestimation of the stenosis diameter compared with manual measurement (Fig 6). 3mensio fits an ellipse on the segmented lumen and

uses the smallest diameter of the ellipse as a minimal stenosis diameter; this can result in a minimal stenosis diameter that is larger than the minimal stenosis diameter measured by a radiologist (Fig 6). This method caused 40% (4/10) of the 3mensio false-negatives for a stenosis of  $\geq 70\%$  and 20.0% (3/15) for a stenosis of  $\geq 50\%$ . For 8.7% (4/46) of the false-negatives for a stenosis of  $\geq 70\%$  and 12.3% (7/57) for a stenosis of  $\geq 50\%$ , the difference in the degree of stenosis with manual measurement was only 5% and the manual measurements were just above the cutoff point and the semiautomatic measurements were just below the cutoff.

### DISCUSSION

In this study, we investigated the diagnostic performance of 4 commercially available semiautomatic software packages with manual measurement as a reference. All semiautomatic methods had a moderate-to-good sensitivity for detecting a stenosis of  $\geq 50\%$  and low sensitivity for detecting a stenosis of  $\geq 70\%$ . All semiautomatic methods had a good specificity and overall diag-



**FIG 4.** Scatterplots of the repeated semiautomatic assessment of the degree of stenosis (percentage) measured by observers 1 and 2.

nostic accuracy for detecting stenoses of  $\geq 70\%$  and  $\geq 50\%$ . All semiautomatic stenosis measurement methods are 40% faster than manual measurements. For 3mensio, we found a much higher interobserver reproducibility compared with manual measurement. All semiautomatic methods had a good correlation with manual measurement.

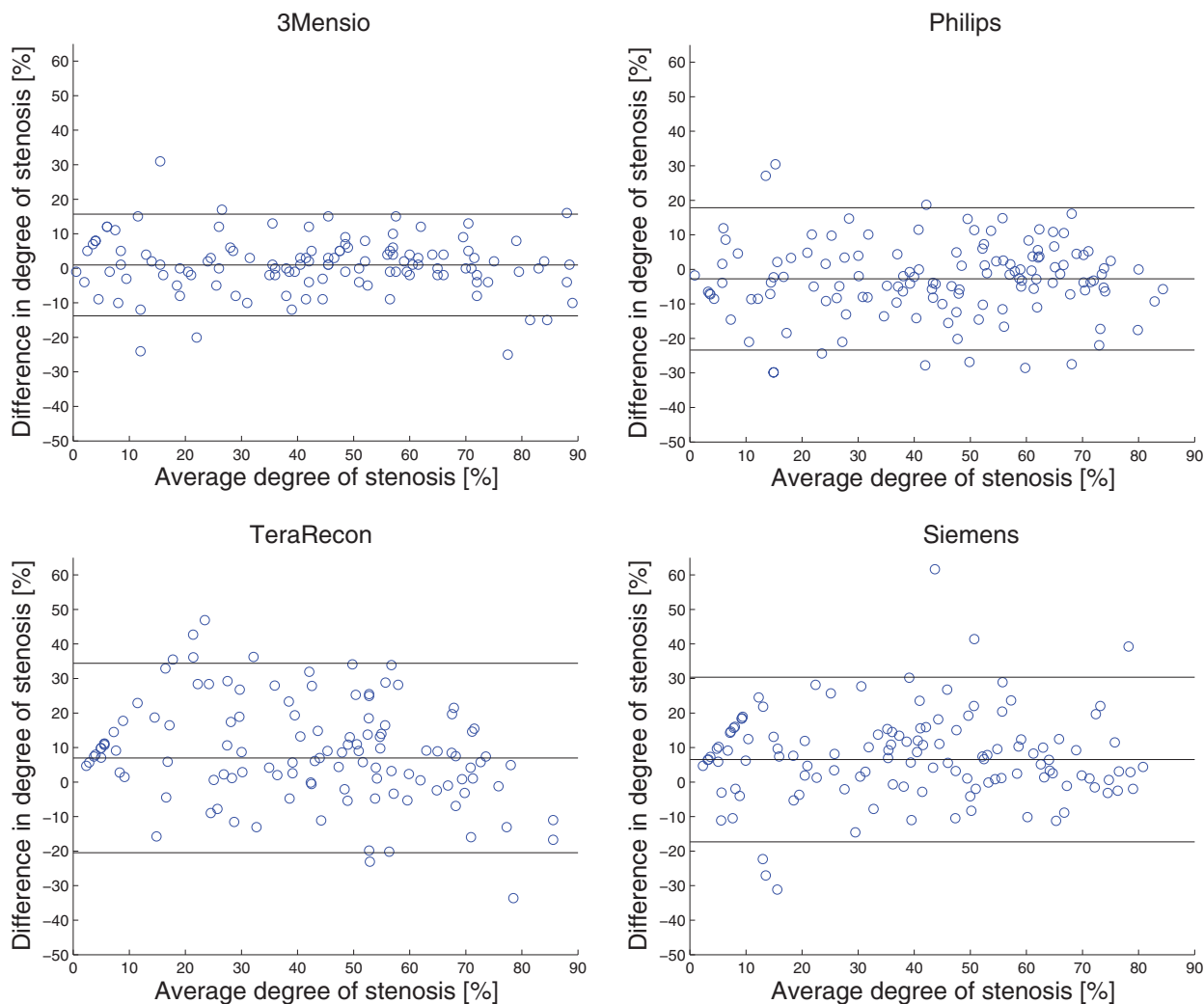
Our results are in line with the previously reported sensitivity and specificity of 75% and 98% for detecting a stenosis of  $\geq 70\%$  and 78% and 93% for detecting a stenosis of  $\geq 50\%$ .<sup>8</sup> Our results are similar to the previously reported sensitivity and specificity of 44.2% and 97.7% for detecting a stenosis of  $\geq 70\%$  and 86.2% and 93.1% for detecting a stenosis of  $\geq 50\%$  in 46 patients with known cerebrovascular disease.<sup>18</sup> The interobserver agreement for semiautomatic measurements is in line with previously reported  $\kappa$  statistics of 0.55 for detecting a stenosis of  $\geq 50\%$ , and 0.59 for detecting a stenosis of  $\geq 70\%$ <sup>19</sup> and Pearson correlation coefficients of 0.89 and 0.90.<sup>9,19</sup> As in previously reported studies,<sup>8,9</sup> we found that semiautomatic stenosis measurement can increase observer reproducibility.

This study has a number of limitations. For pragmatic reasons, we used different observers for different software packages; this difference makes it more difficult to compare the

semiautomatic software packages. We used manual measurements on CTA as a reference, while the original NASCET classification is based on conventional catheter angiography. Due to the risks associated with conventional catheter angiography,<sup>3</sup> it would be unethical to perform DSA. Bucek et al<sup>18</sup> showed that the median difference between semiautomatic CTA and manual DSA stenosis measurement was smaller than the median difference between manual measurement on CTA and DSA,  $-2\%$  versus  $11\%$ , respectively. This finding may imply that manual stenosis measurement tends to overestimate the degree of stenosis compared with measurement on DSA; this overestimation may have caused the low sensitivity found in this study. Due to the low observer reproducibility of manual stenosis measurement, one could question its value as a reference standard to determine the diagnostic accuracy of semiautomatic measurements. However, because manual stenosis measurement is standard in clinical practice, we believed that this measurement was the best choice to evaluate the accuracy of the automated methods.

Retrospective error analysis of the false-negatives showed that most false-negatives were due to the semiautomatic method measuring a larger stenosis diameter and/or a smaller reference diam-





**FIG 5.** Bland-Altman plots of the repeated semiautomatic assessment of the degree of stenosis measured by observers 1 and 2. The *black lines* represent the mean paired difference and 95% limits of agreement.

**Table 4: Observer reproducibility**

	Average Difference Degree of Stenosis $\pm$ SD (%)	Bland-Altman Limits of Agreement (%)	ICC (95% CI) for Degree of Stenosis
Manual intraobserver ( $n = 38$ )	$0.083 \pm 13$	$-26-26$	$0.88 (0.79-0.94)$
Manual interobserver ( $n = 37$ )	$2.6 \pm 15$	$-28-32$	$0.81 (0.70-0.90)$
3mensio interobserver ( $n = 141$ )	$0.94 \pm 7.5^a$ ( $P = .007$ )	$-14-16$	$0.96 (0.95-0.97)^a$ ( $P < .001$ )
Philips interobserver ( $n = 141$ )	$-2.8 \pm 11$	$-23-18$	$0.90 (0.86-0.93)^a$ ( $P = .0041$ )
TeraRecon interobserver ( $n = 141$ )	$7.0 \pm 14$	$-20-34$	$0.83 (0.70-0.90)$
Siemens interobserver ( $n = 141$ )	$6.5 \pm 12$	$-17-30$	$0.86 (0.73-0.92)$

<sup>a</sup> Significant difference with manual interobserver measurements.

**Table 5: Observer reproducibility by statistical  $\kappa$  values**

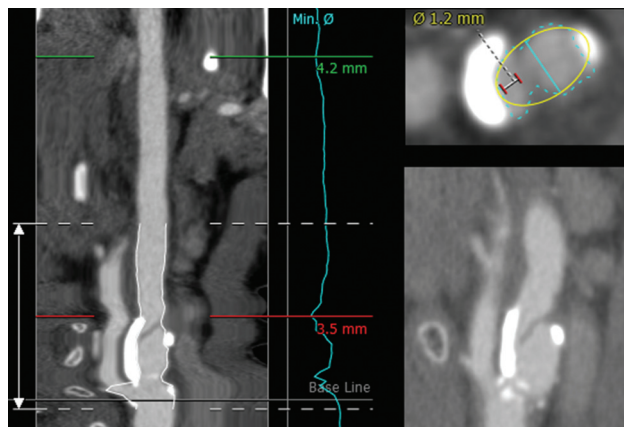
	50% $\kappa^a$ (95% CI)	70% $\kappa^b$ (95% CI)
Manual intraobserver ( $n = 38$ )	$0.73 (0.50-0.95)$	$0.53 (0.02-1)$
Manual interobserver ( $n = 37$ )	$0.73 (0.52-0.95)$	$0.47 (0.03-0.90)$
3mensio interobserver ( $n = 141$ )	$0.88 (0.80-0.96)$	$0.86 (0.73-0.98)$
Philips interobserver ( $n = 141$ )	$0.71 (0.60-0.83)$	$0.53 (0.30-0.77)$
TeraRecon interobserver ( $n = 141$ )	$0.56 (0.41-0.71)$	$0.37 (0.08-0.66)$
Siemens Interobserver ( $n = 141$ )	$0.67 (0.54-0.80)$	$0.63 (0.42-0.84)$

<sup>a</sup>  $\kappa$  values on 50% cutoff.

<sup>b</sup>  $\kappa$  values on 70% cutoff.

eter compared with manual measurements by a radiologist. One-tenth of the false-negatives were caused by an ulcerative plaque that hampered correct semiautomatic measurement and was dif-

ficult for a nonradiologist to detect. To determine the agreement among different pairs of software packages, we randomly selected 1 of the 2 observers for each measurement instead of using the mean of the 2 observers. Averaging the measurements diminishes outliers and therefore might result in a too optimistic agreement between the different semiautomatic software packages.<sup>20</sup> Furthermore in this manner, we aimed to simulate a clinical setting in which multiple users may use the software package.



**FIG 6.** Example of ulcerative plaque. On the *left side* how 3mensio segments the artery is shown, and in the *right upper corner* how 3mensio segments the lumen and the ulcerative plaque is shown. The *turquoise line* is the minimal stenosis diameter as determined by 3mensio (3.5 mm); the *white-with-red line* is a measurement of the true lumen (1.2 mm). The *right lower corner* shows a sagittal view of the ulceration. This image also shows 3mensio fitting an ellipse (yellow) on the segmented lumen of the artery (turquoise).

Creating MPRs perpendicular to the artery, needed for manual stenosis measurement and manual measurement of the diameter of the artery lumen, is prone to observer variation and requires experience.<sup>7,8</sup> This variation resulted in the lower interobserver reproducibility of manual stenosis measurement compared with semiautomatic measurement.

Semiautomatic methods ease stenosis measurements and can have a higher observer reproducibility compared with manual measurement, because manual creation of MPRs and manual lumen measurement are not needed. All 4 semiautomatic software packages are comparable in the ease of use and required observer skills. 3mensio was the only package that determined the minimal diameter of the stenosis automatically. This higher level of automation may have resulted in the superior observer agreement. Although the interobserver reproducibility can be higher for semiautomatic measurements, manual selection of the minimal stenosis diameter and reference diameter is still needed and is therefore a source of observer variability. Furthermore, manual correction of the center line and lumen segmentation are often needed to ensure accurate measurement, especially when the artery is very tortuous or the plaque is calcified.<sup>8,17</sup> The manual selection of the minimal stenosis diameter and reference diameter and the manual corrections may have caused the wide Bland-Altman limits of agreement for the semiautomatic methods and the low observer reproducibility for some of the packages.

Endarterectomy is beneficial for patients with a stenosis degree of  $\geq 50\%$  for men and  $\geq 70\%$  for women.<sup>1</sup> Therefore accurate and reproducible measurement of the degree of stenosis is crucial for selecting patients for endarterectomy.

## CONCLUSIONS

Most semiautomatic software packages have a higher observer reproducibility than manual measurements, which results in more consistent stenosis measurement and less observer dependency in treatment selection. Because of the necessity of manual corrections of semiautomatic measurements, training

of the observers and awareness of erroneous centerlines and lumen segmentations remain crucial. All 4 semiautomatic methods have a high positive predictive value and a good overall diagnostic accuracy for the detection of an ICA stenosis of  $\geq 50\%$  and  $\geq 70\%$ . The potentially excellent observer reproducibility of semiautomatic measurements makes them suitable for clinical practice, but the poor sensitivity for a stenosis of  $\geq 70\%$  should be taken into account and measurements should be checked by a radiologist.

**Disclosures:** Jordi Borst—RELATED: Grant: Information Technology for European Advancement (ITEA)2 project.\* label ITEA 10004: Medical Distributed Utilization of Services & Applications.\* Comments: <https://itea3.org>, <https://itea3.org/project/medusa.html>. Taihra Zadi—UNRELATED: Grants/Grants Pending: GE Healthcare. Paul J. Nederkoorn—UNRELATED: Grants/Grants Pending: 1) The Netherlands Organization for Health Research and Development (No. 171002302), 2) The Netherlands Heart Foundation (No. 2009B095), 3) NutsOhra fund, Comments: 3 grants for investigator-driven research: 1 and 2) The Preventive Antibiotics in Stroke Study, 3) Thrombolysis and Uncontrolled Hypertension study. Rene van den Berg—UNRELATED: Consultancy: agreement with DePuy Codman. Aad van der Lugt—UNRELATED: Grants/Grants Pending: GE Healthcare.\* Comments: MRI vendor and provider of image analysis software; Payment for Lectures (including service on Speakers Bureaus): GE Healthcare.\* Comments: MRI vendor and provider of image-analysis software. Charles B.L.M. Majoie—UNRELATED: Grants/Grants Pending: Dutch Heart Foundation.\* NutsOhra Foundation.\* \*Money paid to the institution.

## REFERENCES

- Rothwell PM, Eliasziw M, Gutnikov SA, et al. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet* 2003;361:107–16
- Barnett HJ, Taylor DW, Eliasziw M, et al. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis: North American Symptomatic Carotid Endarterectomy Trial Collaborators. *N Engl J Med* 1998;339:1415–25
- Willinsky RA, Taylor SM, TerBrugge K, et al. Neurologic complications of cerebral angiography: prospective analysis of 2,899 procedures and review of the literature. *Radiology* 2003;227:522–28
- Koelemay MJW, Nederkoorn PJ, Reitsma JB, et al. Systematic review of computed tomographic angiography for assessment of carotid artery disease. *Stroke* 2004;35:2306–12
- Wardlaw JM, Stevenson MD, Chappell F, et al. Carotid artery imaging for secondary stroke prevention: both imaging modality and rapid access to imaging are important. *Stroke* 2009;40:3511–17
- Bucek RA, Puchner S, Haumer M, et al. Grading of internal carotid artery stenosis: can CTA overcome the confusion? *J Endovasc Ther* 2006;13:443–50
- Howard P, Bartlett ES, Symons SP, et al. Measurement of carotid stenosis on computed tomographic angiography: reliability depends on postprocessing technique. *Can Assoc Radiol J* 2010;61:127–32
- Marquering HA, Nederkoorn PJ, Smagge L, et al. Performance of semiautomatic assessment of carotid artery stenosis on CT angiography: clarification of differences with manual assessment. *AJNR Am J Neuroradiol* 2012;33:747–54
- White JH, Bartlett ES, Bharatha A, et al. Reproducibility of semi-automated measurement of carotid stenosis on CTA. *Can J Neurol Sci* 2010;37:498–503
- Biermann C, Tziflikas I, Thomas C, et al. Evaluation of computer-assisted quantification of carotid artery stenosis. *J Digit Imaging* 2012;25:250–57
- Hameeteman K, Zuluaga MA, Freiman M, et al. Evaluation framework for carotid bifurcation lumen segmentation and stenosis grading. *Med Image Anal* 2011;15:477–88

12. Marquering HA, Majoie CB, Smagge L, et al. **The relation of carotid calcium volume with carotid artery stenosis in symptomatic patients.** *AJNR Am J Neuroradiol* 2011;32:1182–87
13. Marquering HA, Nederkoorn PJ, Bleeker L, et al. **Intracranial carotid artery disease in patients with recent neurological symptoms: high prevalence on CTA.** *Neuroradiology* 2013;55:179–85
14. Bleeker L, Marquering HA, van den Berg R, et al. **Semi-automatic quantitative measurements of intracranial internal carotid artery stenosis and calcification using CT angiography.** *Neuroradiology* 2012;54:919–27
15. Steen WE Van Der, Vermeij J, Marquering HA, et al. **Intracranial carotid artery stenosis diagnosed with CTA in a western population: predictor for poor outcome.** In: Derricks, S ed., *Carotid Artery Disease: Risk Factors, Prognosis and Management. Cardiology Research and Clinical Developments*. Hauppauge, NY: Nova Science Publishers; 2014:33–48
16. Van Dijk AC, Fonville S, Zadi T, et al. **Association between arterial calcifications and nonlacunar and lacunar ischemic strokes.** *Stroke* 2014;45:728–33
17. Bartlett ES, Walters TD, Symons SP, et al. **Quantification of carotid stenosis on CT angiography.** *AJNR Am J Neuroradiol* 2006;27:13–19
18. Bucek RA, Puchner S, Kanitsar A, et al. **Automated CTA quantification of internal carotid artery stenosis: a pilot trial.** *J Endovasc Ther* 2007;14:70–76
19. Zhang Z, Berg MH, Ikonen AEJ, et al. **Carotid artery stenosis: reproducibility of automated 3D CT angiography analysis method.** *Eur Radiol* 2004;14:665–72
20. Maddox WT. **On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization.** *Percept Psychophys* 1999;61:354–74