Reproducibility of Lumen and Vessel Wall Measurements in Carotid Magnetic Resonance Imaging

John A. Roberts*,1, Seong-Eun Kim1, Hyo-Chun Yoon2, Joseph S. McNally1, John R. Hadley1, Laura K. Findeiss1,§, Gerald S. Treiman3 and Dennis L. Parker1

1Department of Radiology, University of Utah, Salt Lake City, UT 84132, USA
2Department of Diagnostic Imaging, Kaiser Permanente, Honolulu, HI 96814, USA
3Department of Surgery, University of Utah, Salt Lake City, UT 84132, USA

Abstract: Atherosclerotic carotid artery disease is estimated to represent the etiology for one quarter of all strokes. Carotid magnetic resonance imaging and magnetic resonance angiography are promising tools in the evaluation of carotid atherosclerotic vascular disease. In this study, we evaluate the reliability of high resolution carotid wall magnetic resonance (MR) imaging by investigating the inter-observer, intra-observer, and inter-scan variability in measurements of carotid vessel total lumen area and mean wall thickness. This HIPAA compliant study received IRB approval and all subjects gave written informed consent. Nineteen subjects were imaged on a 3T MRI scanner with custom-built 4-element receive-only phased-array coils optimized for carotid anatomy. Three observers manually drew regions of interest around the lumen and outer wall for both left and right carotid arteries.

Intraclass correlation coefficients (ICC) showed excellent agreement between Observer 1 and the others (>0.92). A two-way analysis of variance (ANOVA) found no significant difference between observers (p>0.05). Intra-observer variability for Observer 1 was measured by coefficient of variation (CV) with 0.03 for total lumen area and 0.03 for mean wall thickness. Similarly, the inter-scan variability of Observer 1 was found by CV to be 0.05±0.02 for total lumen area and 0.04±0.03 for mean wall thickness. Our results demonstrate that the MR measurements of total lumen area and mean wall thickness are highly reproducible and provide a reliable foundation for the evaluation of carotid atherosclerotic vascular disease.

Keywords: 3 Tesla, MRI, Carotid artery, Reproducibility, Atherosclerosis.

INTRODUCTION

Stroke is the third leading cause of death in the United States, and the leading cause of permanent neurologic disability. Approximately 800,000 Americans will sustain a stroke each year [1], and it is estimated that 25% of these are secondary to atherosclerotic disease in the extracranial carotid arteries. Further, randomized controlled trials including the North American Symptomatic Carotid Endarterectomy Trial, Asymptomatic Carotid Atherosclerosis Study, and others have correlated angiographic stenosis with stroke risk in both symptomatic and asymptomatic individuals [2-5]. The most widely used method of determining stroke risk from carotid stenosis is carotid duplex ultrasonography, combining a grayscale ultrasound image of the carotid wall and lumen, with Doppler signals to measure flow velocity at selected locations in the carotid arteries. Multiple studies have correlated duplex velocity measurements with luminal stenosis as measured by catheter arteriography, so a duplex scan is now an accepted method for assessing degree of luminal occlusion [2-5]. Nevertheless, a duplex scan has many limitations. It can only provide values at selected regions, is dependent on the training and expertise of the technologist, varies with angle of insonation, and is affected by patient body habitus, depth of the vessels, tortuosity, and vessel calcification.

Magnetic resonance imaging (MRI) has been proposed as a method for predicting the risk of neurologic events secondary to carotid stenosis [6-11]. MRI is characterized by a high tissue contrast, with the ability to distinguish different tissues from one another based on their differences in MRI properties such as T1 and T2 relaxation times. Developments in hardware and image acquisition pulse sequence design now allow for detailed identification and measurement of the components of carotid plaque including hemorrhage, fibrous cap, lipid/necrotic core and active inflammation that have been related to cerebral embolization.

Although MRI has the potential to accurately identify plaque components and lumen stenosis, concerns over scan reliability and measurement consistency limit its widespread use in clinical practice. Consequently, several groups have studied inter-scan, intra-observer and inter-observer variability in the MRI of carotid atherosclerosis. Wall and lumen size measurements exhibit the highest consistency with inter-scan intraclass correlation coefficients (ICC) ranging from 0.79 [12] to 0.99 [13-19], and coefficients of variation (CV) ranging from 0.026 [13] to 0.089 [15]. Intra-
and inter-observer measurements of the wall and lumen have associated ICC values ranging from 0.96 [12, 16, 19] to 0.99 [17] and CV values ranging from 0.05 to 0.18 [12]. The measurements of carotid plaque components such as lipid-rich/necrotic core, calcification and hemorrhage tend to exhibit more variability with ICC values ranging from 0.70 [17] for hemorrhage volume to 0.96 [18] for calcium volume and CV values ranging from 0.11 [14] to 0.86 [12] for lipid volume.

Since the publication of our previous work [20], our research platform was upgraded from a 1.5T to a 3T MRI system. We performed measurements of reproducibility to study the effects of the new hardware and MRI sequences optimized for the higher field.

MATERIALS AND METHODOLOGY

Study Subjects

This HIPAA compliant study received IRB approval and all subjects gave written informed consent. Subjects recruited for this study were divided into two groups. Nine subjects (Group 1) were imaged once. Group 1 images served as both the training set for the primary observer (Observer 1) and as the basis of an inter-observer study. Ten additional subjects (Group 2) were imaged three times in succession. Intra-observer and inter-scan studies of Group 2 images were conducted by Observer 1.

Imaging

All subjects were imaged on a Siemens TIM Trio 3T MRI scanner (Siemens Medical Solutions, Erlangen, Germany) with custom-built 4-element receive-only phased-array coils optimized for carotid anatomy [21, 22]. The subjects were positioned comfortably in a custom-built head/neck holder described previously [20] (Fig. 1). The head/neck holder was employed simply to reduce motion and was not used in its capacity as a repositioning device. In this study, the post and nose bridge (D) were not used and all subjects were imaged with the same Siemens foam head rest (not shown).

Group 1 Subjects

Three series of images were acquired for this study: a rapid 2D Time of Flight (TOF), a 3D TOF and a 2D T1-weighted (T1w) turbo spin echo with our modified version of the double inversion preparation (DIR) [23]. All images were acquired without electrocardiography gating. The carotid bifurcation and flow dividers were localized with 2D TOF. 3D TOF images were acquired at a location centered approximately with the flow dividers along the axial direction. 2D DIR T1w images were acquired centered axially with the flow dividers. The blood signal in T1w images was suppressed using DIR with inversion time of 500ms. A chemical-shift fat saturation RF pulse was applied to eliminate perivascular fat. Both the 3D TOF and 2D DIR T1w images were acquired with an inplane resolution of 0.5×0.5 mm² and displayed with resolution 0.25×0.25 mm² following zero filled interpolation. Table 1 summarizes the remaining imaging parameters used.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>3D TOF</th>
<th>2D DIR T1w</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>repetition time (TR)</td>
<td>30</td>
<td>800</td>
<td>ms</td>
</tr>
<tr>
<td>echo time (TE)</td>
<td>3.86</td>
<td>8.8</td>
<td>ms</td>
</tr>
<tr>
<td>flip angle</td>
<td>25</td>
<td>degrees</td>
<td></td>
</tr>
<tr>
<td>number of averages (NEX)</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>echo train length (ETL)</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>field-of-view (FOV)</td>
<td>130</td>
<td>130</td>
<td>mm</td>
</tr>
<tr>
<td>receiver bandwidth (BW)</td>
<td>180</td>
<td>250</td>
<td>Hz/pixel</td>
</tr>
<tr>
<td>number of slabs</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of slices</td>
<td>72¹</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>slab overlap</td>
<td>26</td>
<td>slices</td>
<td></td>
</tr>
<tr>
<td>slice thickness</td>
<td>0.6</td>
<td>2.0</td>
<td>mm</td>
</tr>
</tbody>
</table>

¹Per slab.

Group 2 Subjects

The imaging protocol used for Group 1 subjects was also used for Group 2 subjects. For each subject, the same protocol was repeated three times over the course of 90 minutes. The three imaging scans were acquired consecutively for subject convenience and to limit any confounding effects due to disease progression. To approximate inter-scan variability due to changes in patient and coil positioning, each subject was completely removed from the scanner between repetitions of the imaging protocol. The subjects were asked to stand briefly and then were repositioned on the imaging bed with the head/neck holder device. As noted, the head/neck holder was used to reduce motion during a scan but not to improve repositioning between scans. We assume that for carotid imaging in a clinical setting, intra-scan immobilization devices are more
often available and used in practice than inter-scan repositioning devices.

**Contour Definition**

For all T1w images, the common carotid artery below the flow divider and internal carotid artery above the flow divider were outlined manually using software we developed in IDL 7.0 (ITT Visual Information Solutions, Boulder, CO). Contours were drawn by three observers for the lumen and outer vessel wall on each of the 24 T1w images for both the left and right carotid arteries. Contours were drawn on the common carotid artery below the flow divider and the internal carotid artery above the flow divider. A total of 96 contours were drawn per scan per subject per observer. The TOF images were aligned axially with the T1w images based on flow divider location and displayed in a separate window as an aid in the interpretation of the T1w vessel lumen and wall. Contours were drawn in order from the most inferior slice to the most superior slice: first, for all lumens; then, for all outer vessel walls; and finally, the same order was repeated for the contralateral carotid. Thus, lumen contours were visible during the drawing of the outer vessel wall. However, because the left and right sides were drawn separately, contours of the contralateral carotid were not visible.

**Lumen Area and Wall Thickness**

A single lumen area and wall thickness value was computed per subject per experiment. Though contours were drawn for both arteries in all 24 T1w images for each subject, only a subset of images was used to compute a lumen area and wall thickness value. This subset, which was consistent in each of the variability experiments described below, was selected so that roughly the same vascular anatomy was measured along the axial direction. The subset always included the bifurcation location, with the number of slices proximal and distal to the bifurcation varying due to relative registration between subjects or exams. Separate subsets were selected for the left and right sides so that the same anatomy was measured relative to the respective bifurcations.

Total lumen area and total wall area were computed by summing contour areas over the subset of images, effectively computing a volume. The values for the left and right sides were summed per subject to produce a single lumen and outer wall area per subject per experiment. Though in practice measurements of the left and right arteries would be considered separately, we sum them here for purposes of assessing repeatability.

A mean contour radius, \( r \), was determined by equating each total area (TA) measure to that of a cylinder of length \( 2N \):

\[
\bar{r} = \sqrt{\frac{TA}{2N\pi}}
\]

where \( N \) was the number of images considered per side in the calculation of TA. It is acknowledged that one cannot in general assume a circular cross section for either the inner lumen or outer vessel wall. However, for these experiments the regions considered were centered on the bifurcation and roughly cylindrical. One might improve the estimate of the mean radius by substituting the centerline length of vessel segment for \( N \).

A mean wall thickness, MWT, was computed by comparing the mean lumen radius, \( \bar{r}_{lumen} \), with the mean outer wall radius, \( \bar{r}_{outer} \):

\[
MWT = \bar{r}_{outer} - \bar{r}_{lumen}
\]

**Inter-Observer Variability**

The images of the Group 1 were used as a training set to compare measurements made by the non-clinician, Observer 1, with measurements made by a 3rd year radiology resident, Observer 2, and a board certified radiologist with over 10 years experience in reading vascular imaging studies, Observer 3. For each pair of observers, Bland-Altman graphs were constructed to assess bias and error [24]. Paired and three-way ICC values were computed. A two-way analysis of variance (ANOVA) was used to assess the relative contributions to variability by subject and observer for both the total lumen area and mean wall thickness.

**Intra-Observer Variability**

The images of one scan for one subject from Group 2 were selected for the purpose of measuring intra-observer variability. For these images, Observer 1 drew complete sets of contours repeatedly on ten separate occasions. The coefficient of variation,

\[
CV = \frac{\sigma}{\mu}
\]

was computed for total lumen area and mean wall thickness.

**Inter-Scan Variability**

Total lumen area and mean wall thickness were computed for each of the three scans for each subject in Group 2. Population mean and standard deviations were estimated from these three scan measurements and a single CV computed per subject. A global mean, \( \mu_{CV} \), and standard deviation, \( \sigma_{cv} \), was computed from the individual subject CV values. Agreement between scans was measured with a three-way ICC. A two-way analysis of variance (ANOVA) was used to assess the relative contributions to variability by subject and scan for both the total lumen area and mean wall thickness.

**RESULTS**

**Inter-Observer Variability**

Fig. (2) shows representative TOF and T1w images and contours for one subject in the inter-observer study. Eight consecutive T1w images covering 1.6 cm axially beginning at the left carotid bifurcation (left column) are shown in the second row with the nearest TOF images shown in the first row. The T1w images are repeated in rows 3 through 5 with the contours drawn by Observers 1, 2, and 3 respectively. For comparison, the agreement between observers was lower for the subject of Fig. (3). As in Fig. (2), the images extend axially 1.6 cm above the left carotid bifurcation. There is some disagreement regarding the contour of the inner lumen.
for the second and third column of images. Arrows indicate possible flow artifacts in the original images which may have influenced image interpretation. Additionally, Observer 2 drew a somewhat smaller outer wall contour and hence produced a lower estimate of wall thickness.

Fig. (4) shows representative Bland-Altman graphs comparing total lumen area measurements between Observer 1 and the other two observers for each of 9 subjects. For these subjects, the common anatomical region consisted 8 slices or 1.6 cm superior to the bifurcation. Graphs for mean wall thickness and those comparing Observer 2 and Observer 3 were analyzed but are not shown here. Observer 1 exhibits a minor bias on the order of 0-4% of the mean of the measured lumen areas with respect to Observer 2. Observer 1 can be expected to differ by less than 17% of the mean from the values measured by the other observers 95% of the time. A two-way ANOVA indicated that for mean values of the total lumen area and mean wall thickness, intra-subject differences were significant with $F(8,16)=174$ ($p<0.01$) and $F(8,16)=48.6$ ($p<0.01$), respectively. In contrast, inter-observer means were not significantly different with $F(2,16)=1.6$ ($p>0.2$) for total lumen area and $F(2,16)=2.9$ ($p>0.05$) for mean wall thickness. Table 2 summarizes the results from the Bland-Altman and ICC analysis.

**Intra-Observer Variability**

In this study, the same anatomical region was used for analysis as for the inter-observer study: 8 slices or 1.6 cm axial coverage beginning at the bifurcation. The CV values for total lumen area and mean wall thickness were both 0.03. That is, Observer 1 shows roughly a 3% variation about the mean when measuring total lumen area or mean wall thickness.
Reproducibility of Lumen and Vessel Wall Measurements

In the intra-observer comparison, the range of areas and wall thicknesses was smaller than that of the inter-observer comparison.

Inter-Scan Variability

As with the inter- and intra-observer studies, in this analysis a 1.6 cm volume was selected above each bifurcation for measurement. A Bland-Altman analysis of the total lumen area and mean wall thickness without regard to subject showed less than 1% inter-scan bias. Under the assumption of zero bias, Fig. (5) shows the variability in total lumen area and mean wall thickness as the difference between measurements and the mean versus the mean. No dependence of the variability on the size of the measurements is apparent. The CV was computed for each of the ten subjects analyzed and a population mean CV determined. Data ranges, population mean CV, and ICC values are shown in Table 3.

Table 2. Inter-Observer Variability Measurements

<table>
<thead>
<tr>
<th>Observers</th>
<th>Total Lumen Area</th>
<th>Mean Wall Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range (mm²)</td>
<td>Bias (mm²)</td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>400±172</td>
<td>-15.7</td>
</tr>
<tr>
<td>1 &amp; 3</td>
<td>392±171</td>
<td>1.6</td>
</tr>
<tr>
<td>2 &amp; 3</td>
<td>400±171</td>
<td>17.3</td>
</tr>
<tr>
<td>all</td>
<td>398±166</td>
<td>0.988</td>
</tr>
</tbody>
</table>

Ranges represent mean and standard deviation of all values compared. Bias values and agreement intervals (1.96σ) were taken from the Bland-Altman analysis.

Variability as measured by CV increased for the inter-scan case as compared to the intra-observer study. A Student t-test analysis showed a significant difference between the intra-observer and inter-scan means for total lumen area CV (p<0.01) and mean wall thickness CV (p<0.05). One expects a larger mean CV for the mean wall thickness due to its dependence on two observer measurements rather than one. Inter-scan agreement as measured by ICC was high. The two-way ANOVA showed strong support for the null hypothesis that inter-scan means were the same for either total lumen area, F(2,18)=0.1 (p>0.9) or mean wall thickness, F(2,18)=0.54 (p>0.5).

DISCUSSION

Inter-Observer Variability

Based on ICC values >0.9, the results of Observer 1 show good agreement with those of the trained observers. Because mean wall thickness is computed as the difference of two measurements, we expect variability to increase and correlation between observers to decrease. Our three-way ICC of 0.980 for total lumen area compares favorably with the inter-observer ICC values of 0.998 reported by Dehnavi et al., [16], 0.98 reported by Touzé et al., [17] and 0.927-0.991 reported by Syed et al., [19]. In light of the two-way ANOVA which found no significant difference between observers (p>0.05), we concluded that the 4% relative bias...
found in the Bland-Altman analysis was not significant. For purposes of the inter-scan study, we concluded that the agreement and near zero bias between Observer 1 and Observer 3, the senior physician, was sufficient to validate the performance of Observer 1.

Intra-Observer Variability

Our CV of 0.03 for intra-observer variability for the total lumen area was lower than those reported for studies performed at 3T by Yarnykh et al., [15] (0.071) and Dehnavi et al., [16] (mean relative error, 0.048), but essentially equal to the variability reported by Vidal et al., (0.026). Similarly, our CV of 0.03 for mean wall thickness was less than that of Yarnykh et al., (0.098). The CV of 0.03 provides an important baseline for the interpretation of the inter-scan variability. It is possible to compute an estimate of percent stenosis using either areas or mean diameters of the lumen contours. While stenosis was not computed in this study, the inter- and intra-observer results suggest that MRI of the carotid anatomy is of sufficient quality that a trained non-radiologist could draw contours and arrive at stenosis estimates that are very close to those of trained radiologists. This has implications for workflow in future large scale studies when one considers that a technologist or resident could draw contours and arrive at a measure of stenosis requiring only review by a radiologist or vascular surgeon.

Inter-Scan Variability

Our mean CV inter-scan variability measure for total lumen area at 3T, 0.05, falls between the values of 0.026 and 0.086 reported by Feiyu et al., [18] and Vidal et al., [13], respectively. Similarly, the ICC value of 0.956 is lower than the 0.99 reported by Feiyu et al., and 0.982 reported by Dehnavi et al., but essentially equal to the 0.95 reported by Vidal et al. Our mean CV variability for mean wall thickness at 3T, 0.04, compares favorably with the 0.0387 reported by Feiyu et al. However, the corresponding ICC value of 0.906 is lower than the 0.97 reported by Feiyu et al. The use of automated analysis software by Feiyu et al., to extract lumen and wall contours likely explains their improved variability scores and high inter-scan ICC values.

The inter-scan variability measured for both total lumen area and mean wall thickness was less than twice that of the intra-observer variability. Because Observer 1 conducted both these tests, the results indicate that over half the measured inter-scan variability might be accounted for by inherent observer variability with the remainder due to true inter-scan variability. This is consistent with the observation by Wasserman et al., who concluded that intra-observer variability was the main source of variability in the quantitative measurements [12]. The relative bias between observers as discussed above only increases the dependence of variability upon the observer. One might reduce the bias between observers by further training and the establishment of standards for the manual drawing of contours.

Measurement and Analysis

Both how the measurements are made and how the analysis is performed have direct effects upon the resulting estimates of variability and agreement. For the intra-observer study, our CV estimates were based on ten repeated measurements of one exam of a single subject while those of Yarnykh et al., were based on the root-mean-square combination of CV values for 7 subjects. Similarly, Dehnavi et al., based their estimate on 10 subjects. It is possible that their higher variability values reflect some inter-subject variability in addition to intra-observer variability. However, it is more likely that the observed differences in reported variability are related to the size of the underlying contours used to perform the measurements.

The spatial extent measured has an important effect upon the resultant estimate of variability. The spatial extent measured by Vidal et al., per subject was greater than our own, encompassing the common, internal and external carotids on both sides. The measurements of Feiyu et al., had twice the axial coverage of our measurements, 3.2 cm compared to 1.6 cm. Dehnavi et al., measured only the left side for their subjects whereas our measurements combined both sides for each test subject. Yarnykh et al., who measured areas rather than volumes, make a convincing argument that much of the observed variability arises from pixel-sized variations in the contour definition [15]. It follows that measurement variability will decrease with an increase in the total number of pixels or the spatial extent underlying a given measurement. Measurements made in diseased populations may show lower variability simply because diseased vessel walls are larger than those of normal populations. It is likely that our summing over the left and right sides of each subject reduced variability in our measurements. The tradeoff for lower measurement variability is a decrease in sensitivity to small but real changes in inter-scan comparisons.

Limitations of this Study

The goal of this study was to assess the reproducibility of carotid artery measurements using the latest MR hardware and image analysis tools. This study neither supports the replacement of carotid duplex ultrasonography in particular or the use of MRI in general for the clinical treatment of carotid artery disease. Establishing the clinical relevance of MRI in the treatment of carotid artery disease will require much greater efforts, enrolling hundreds if not thousands of subjects. However, this reproducibility study and other cited reproducibility studies establish confidence in the underlying MRI measurements. Though the number of subjects in this study was small, we believe that our experimental and statistical methodologies support the resulting quantitative assessment of reproducibility.

CONCLUSION

The observed low inter-scan variability supports the continued assessment of MRI in future studies of carotid artery disease. For examples of past and ongoing studies of MRI in the clinical treatment of carotid artery disease, the reader is directed to references [10, 12, 25-27]. The lower the inter-scan variability, the smaller the degree of disease progression one can confidently measure and compare with results from prevailing clinical tools such as duplex ultrasonography, computed tomography angiography, or contrast arteriography. We anticipate that the increased use of automated software analysis tools will drive the further reduction of measurement variability as well as begin to eliminate inter-observer bias. This study suggests that high resolution 3T MRI of the carotid arteries using dedicated
surface coils provides a very precise estimate of carotid artery luminal area and wall thickness.

ACKNOWLEDGEMENTS

This work was supported in part by: a grant from the National Institutes of Health (HL 57990); a merit review grant from the Department of Veterans Affairs; the Ben B. and Iris M. Margolis Foundation; Siemens Medical Solutions; and the Mark H. Huntsman Endowed chair.

We would like to thank: our volunteer coordinator, Anne Haroldsen, for her work in recruiting subjects; our coil engineer, J. Rock Hadley, and members of his lab for our custom designed carotid phased-array coils; and our MR technologists, Melody Johnson and Henry Buswell, for their assistance in imaging the volunteers.

REFERENCES


Received: September 21, 2011 Revised: November 3, 2011 Accepted: November 4, 2011

© Roberts et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.