Validation of Automatically Classified Magnetic Resonance Images for Carotid Plaque Compositional Analysis

Sharon E. Clarke, MD, PhD; Vadim Beletsky, MD; Robert R. Hammond, MD; Robert A. Hegele, MD; Brian K. Rutt, PhD

Background and Purpose—MRI may be used for noninvasive assessment of atherosclerotic lesions; however, MRI evaluation of plaque composition requires validation against an accepted reference standard, such as the American Heart Association (AHA) lesion grade, defined by histopathological examination.

Methods—Forty-eight carotid endarterectomy specimen cross-sections had AHA lesion grade determined histopathologically and were concurrently imaged using combinations of 8 MRI contrast weightings in vitro. A maximum likelihood classification algorithm generated MRI “maps” of plaque components, and an AHA lesion grade was assigned correspondingly. Additional analyses compared classification accuracy obtained with a commonly used set of magnetic resonance contrast weightings [proton density (PDw), T1 (T1w), and partial T2 (T2w)] to accuracy obtained with the combination of PDw, T1w, and diffusion-weighted (Dw) contrast.

Results—For the 8-contrast combination, the sensitivities for fibrous tissue, necrotic core, calcification, and hemorrhage detection were 83%, 67%, 86%, and 77%, respectively. The corresponding specificities were 81%, 78%, 99%, and 97%. Good agreement (79%) between magnetic resonance and histopathology for AHA classification was achieved. For the PDw, T1w, and Dw combination, the overall classification accuracy was insignificantly different at 78%, whereas the overall classification accuracy using PDw, T1w, and partial T2w contrast weightings was significantly lower at 67%.

Conclusions—This study provides proof-of-principle that the composition of atherosclerotic plaques determined by automated classification of high-resolution ex vivo MRI accurately reflects lesion composition defined by histopathological examination. (Stroke. 2006;37:93-97.)

Key Words: atherosclerosis ■ carotid arteries ■ MRI
Methods

General Strategy

Carotid endarterectomy specimens were obtained at surgery and imaged with both MRI and microcomputed tomography (μCT). All of the investigators were blinded to any clinical or personal data of the operated patients. In particular, to avoid selection bias, special care was taken to obscure information on whether patients were symptomatic and, if yes, which side had been involved. Also, laboratory data were not recorded, ensuring that investigators were not aware of lipid profiles, homocysteine levels, or any statin agents administered.

After imaging, histology was obtained; well-matched axial MR, μCT, and histological images were selected; and the histological and μCT images were registered to the corresponding MR images. The MR images were classified on a pixel-by-pixel basis into 4 tissue classes: fibrous/loose connective tissue, hemorrhage, necrosis, and calcification, using a maximum likelihood classification algorithm. The accuracy of the classified images was determined by pixel-by-pixel comparison to a pathologist’s interpretation of the histopathology. AHA grading was derived both conventionally from the histological images and from the classified MR images, and these gradings were compared.

MR and μCT Imaging

MRI and μCT imaging of the endarterectomy specimens were performed using a protocol described previously.8 Image acquisition parameters were as follows: proton density weighted with repetition time (TR)/echo time (TE) 2000/7.8 ms; partial T2w with TR/TE 2000/32 ms; T2w with TR/TE 2000/62 ms; T1w with TR/TE 500/7.8 ms fast spin-echo; fast imaging using steady-state acquisition with TR/TE/flip 5.3/2.5/35; T1w spoiled gradient echo with TR/TE/flip 35/5.9/25; T1w spoiled gradient echo with magnetization transfer, TR/TE/flip 35/5.9/25; and Dw (b=1500 s/mm2) spin-echo with TR/TE 2000/25 ms. The in-plane resolution for all of the sequences was 156×156 μm, and the slice thickness ranged from 200 to 1000 μm.

After MRI, the endarterectomy samples were fixed in formalin and imaged with μCT.8 Because the specimens were decalcified before histological sectioning, the μCT images served as the gold standard for identification of calcified plaque regions.

After imaging, 4-μm-thick sections were taken at 0.5-mm intervals along the length of the plaque and stained with Movat’s pentachrome and hematoxylin/eosin. The histological sections provided the reference standard for the presence of various plaque components other than calcification.

MR Image Processing and Classification

From the 13 carotid endarterectomy specimens, 48 axial cross-sections were selected for analysis, with a range of 3 to 5 slices per specimen. Although there were ≥48 cross-sections available, MR, μCT, and histological images acquired, these 48 were selected as the minimal set for which all 3 of the image types were available, artifact-free, registerable to one another, and subject to the requirement that within 1 specimen, selected cross-sections were not too similar (ie, different total plaque cross-sectional area, arrangement, and/or number of plaque components). To avoid selection bias, slices were selected before automated classification and pathologist interpretation. Selection based on these criteria maximized the range of plaque and tissue types included in the study while minimizing errors in classification because of registration inaccuracies.

The MR images of the endarterectomy specimens were classified using the maximum likelihood classifier, a commonly used supervised classification algorithm.10 Small regions of interest representing fibrous/loose connective tissue, lipid-rich necrosis, calcification, and hemorrhage were selected from the 12 endarterectomy slices in the training data set. These training regions constituted only 2.5% of the total plaque cross-sectional area and were selected by an expert pathologist to enclose regions of unambiguously defined, uniform tissue on the basis of examination of the histological images. These small training regions were selected from 12 of the 48 cross-sections, spanning and uniformly distributed across 8 of the 13 endarterectomy specimens (the “training set”). The fact that such a small proportion of the total plaque pixels was used for training meant that the entire set of 48 cross-sections (the “test set”) could be used to independently validate the classifier (97.5% of the analyzed pixels were different than the training pixels).

Based on the training data, the signal intensity distribution for each plaque component was modeled as a multivariate Gaussian distribution. Initial analysis showed that the pixel percentage and training region distribution we used produced a robust estimate of the underlying intensity distributions (mean intensity and SD, both used for pixel classification with this method). For every pixel in the test set, the maximum likelihood classification algorithm reports the tissue class with the highest probability given the overlapping distributions modeled by the training data. After the classifier was trained, all 48 of the cross-sectional plaques were fully classified.

Validation of Classified MR Images by Comparison With Histopathology

The accuracy of the classifier was determined by comparing the classified image on a pixel-by-pixel basis with μCT and histology, which served as the gold standards for calcification and other plaque components, respectively. The μCT images were registered to the MR images using linear registration (scaling, rotation, and translation only), with the ImageJ plug-in Align3TP software (National Institutes of Health), whereas the digitized histological images were registered to the MR images using the Delaunay triangle nonlinear registration method (ENVI, Research Systems Incorporated) so that shearing and other distortions caused by histological processing could be corrected. Applying a threshold to the μCT images generated “truth” regions of interest for calcification. A pathologist’s interpretation of the histology served as the truth regions for other plaque components.

### TABLE 1. Modified AHA Classification

<table>
<thead>
<tr>
<th>Modified AHA Lesion Type</th>
<th>Description of Criteria for Classified Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I-II</td>
<td>Near normal wall thickness</td>
</tr>
<tr>
<td>Initial lesion/fatty streak</td>
<td>No calcification</td>
</tr>
<tr>
<td>with foam cells</td>
<td>No lipid core</td>
</tr>
<tr>
<td>Type III</td>
<td>Small plaque with minimal luminal narrowing</td>
</tr>
<tr>
<td>Preatheroma with</td>
<td>Small lipid-rich or necrotic core, &gt;15% of plaque</td>
</tr>
<tr>
<td>extracellular lipid pools</td>
<td>No calcification</td>
</tr>
<tr>
<td>Type IV-Va</td>
<td>Well-defined lipid pool or necrotic core, &gt;15% of plaque</td>
</tr>
<tr>
<td>Atheroma with</td>
<td>Intact fibrous cap</td>
</tr>
<tr>
<td>confluent lipid core</td>
<td>Calcification may be present</td>
</tr>
<tr>
<td>Type VI</td>
<td>No hemorrhage</td>
</tr>
<tr>
<td>Complicated plaque</td>
<td>Complicated plaque with hemorrhage or rupture</td>
</tr>
<tr>
<td></td>
<td>Positive identification of hemorrhage was &gt;150 contiguous pixels* labeled hemorrhage</td>
</tr>
<tr>
<td></td>
<td>Positive identification of plaque rupture was ≥2 necrotic pixels* communicating with the lumen, which were connected to a pool of necrosis of &gt;35 pixels*</td>
</tr>
<tr>
<td>Type Vb</td>
<td>&gt;15% of area labeled calcium</td>
</tr>
<tr>
<td>Calcified plaque</td>
<td>Necrotic content &lt;15% area of plaque</td>
</tr>
<tr>
<td></td>
<td>No hemorrhage or plaque rupture</td>
</tr>
<tr>
<td>Type Vc</td>
<td>&gt;75% area labeled fibrous</td>
</tr>
<tr>
<td>Fibrous plaque</td>
<td>Necrotic content &lt;15% area of plaque</td>
</tr>
<tr>
<td></td>
<td>No hemorrhage or plaque rupture</td>
</tr>
</tbody>
</table>

*1 pixel was 78×78 μm².
Using an electronic tablet, the pathologist, blinded to both the classifier results and the appearance of the MR images, manually traced appropriate regions of interest onto the digitized histological images, classifying all of the pixels within the plaque into 1 of the 4 defined tissue classes.

Registration of histology to the MR images permitted the 2 images to be overlaid for direct comparison of the MR-derived classification of each individual pixel to the truth class for that same pixel. Thus, the registration of histology to MR allowed for the sensitivity, specificity, and overall accuracy to be calculated for each plaque component on a pixel-by-pixel basis.

Evaluation of Sensitivity, Specificity, and Accuracy of MR Classification

The sensitivity, specificity, and overall accuracy for all of the plaque components were determined for the following 3 combinations of MR contrast used in the classification process: (1) all 8 contrast weightings; (2) only PDw, T1w, and Dw contrast weightings; and (3) only PDw, T1w, and partial T2w contrast weightings. This set has often been used in other studies.

The overall accuracies for each of the 3 conditions were compared with a paired t test. For this purpose, data from cross-sections extracted from the same endarterectomy specimen were combined to ensure that the t statistic was calculated using truly independent samples (n=13 endarterectomy specimens), a requirement for the t test.13

Modified AHA Grading System for Use With Classified MR Images

The AHA scientific statement described the characteristic components and pathogenic mechanisms of early and advanced atherosclerotic lesions. Because MRI is not able to resolve tissue structure at the histological level, the AHA classification criteria were modified for use with noninvasive image data. This was accomplished by combining type I and II lesions, as well as type IV and Va lesions, and using additional criteria as proposed previously.11,14 Table 1 details the specific criteria used to define this modified AHA grading system, which were applied to the MR-generated classification maps.

Results

Intrarater Variability

The intrarater variability measurement for our pathologist-defined classifications yielded overall agreement of 96.8% (κ=0.941) between repeated classifications, separated by 6 weeks. For practical reasons, this study did not involve multiple pathologists, and, therefore, intrarater variability was not assessed.

Accuracy of Maximum Likelihood Classification of MR Images

The sensitivities and specificities for each plaque component were calculated for the 3 different conditions (Table 2). A paired t test demonstrated that the overall accuracy obtained with the best set of 3 contrast weightings (PDw, T1w, and...

<table>
<thead>
<tr>
<th>Variable</th>
<th>All 8 Contrasts</th>
<th>PDw, T1w, Dw</th>
<th>PDw, T1w, Partial T2w</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Accuracy</td>
<td>78±15%*</td>
<td>78±15%*</td>
<td>67±20%</td>
</tr>
<tr>
<td>Component</td>
<td>Sens Spec</td>
<td>Sens Spec</td>
<td>Sens Spec</td>
</tr>
<tr>
<td>Fibrous/loose connective tissue</td>
<td>83 81</td>
<td>80 85</td>
<td>74 68</td>
</tr>
<tr>
<td>Necrotic</td>
<td>67 78</td>
<td>68 78</td>
<td>46 70</td>
</tr>
<tr>
<td>Calcification</td>
<td>86 99</td>
<td>86 99</td>
<td>83 99</td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>77 97</td>
<td>91 95</td>
<td>81 97</td>
</tr>
</tbody>
</table>

*Significantly different from overall accuracy observed for case 3. †loose connective tissue.

Table 3. AHA Classification by MR and Pathology

<table>
<thead>
<tr>
<th>MR Classification</th>
<th>III</th>
<th>IV-Va</th>
<th>VI</th>
<th>Vb</th>
<th>Vc</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>III</td>
<td>3</td>
<td>11</td>
<td>5</td>
<td>13</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>IV-Va</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>17</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vb</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vc</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>15</td>
<td>19</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Percentage agreement, 79%.

For example, for the classified MR images, a plaque was labeled type VI if either the labeled necrotic core (of size >35 pixels) communicated directly with the lumen at any point (with contact length of at least 2 pixels) or a contiguous region of >150 pixels was labeled hemorrhage.

The corresponding histological sections were also classified according to AHA guidelines. This grading was performed by an expert pathologist using the traced and fully classified histological images. The AHA grade assigned to the classified MR image was compared with the true AHA grade as defined by the pathologist.

Results

Intrarater Variability

The intrarater variability measurement for our pathologist-defined classifications yielded overall agreement of 96.8% (κ=0.941) between repeated classifications, separated by 6 weeks. For practical reasons, this study did not involve multiple pathologists, and, therefore, intrarater variability was not assessed.

Accuracy of Maximum Likelihood Classification of MR Images

The sensitivities and specificities for each plaque component were calculated for the 3 different conditions (Table 2). A paired t test demonstrated that the overall accuracy obtained with the best set of 3 contrast weightings (PDw, T1w, and...
Dw) (78%) was not significantly different from that obtained using all 8 of the contrast weightings (78%; \( P = 0.42 \)). In addition, the classifier accuracies achieved with all 8 of the contrasts and with these best 3 contrast weightings were both significantly higher than the accuracy attained with a commonly used combination of 3 contrasts (PDw, T1w, and partial T2w; 67%; \( P = 0.004 \) and 0.010, respectively).

Comparison of AHA Plaque Classification by MR and Histopathology

Our sample of endarterectomy specimens consisted primarily of advanced atherosclerotic lesions (type IV-Va and type VI). Table 3 summarizes the results of AHA grading derived from maximum likelihood classification based on an input of 8 MR contrast weightings and AHA grading based on pathological classification for all 48 of the endarterectomy cross-sections analyzed in this study. The overall agreement was 79\% (\( \kappa = 0.711 \)). Figures 1 and 2 illustrate the different categories of AHA classification.

Often, classified images of the same plaque with a different set of contrast weightings would be assigned the same AHA rating, although the pixel-by-pixel accuracies differed. For example, Figure 3 demonstrates a type IV-Va plaque as assigned by histological examination. Maximum likelihood classification with PDw, T1w, and Dw agreed well with the pathologist’s tracing of the histology (pixel-by-pixel accuracy of 78\%) and also correctly graded the plaque as type IV-Va. Classification based on PDw, T1w, and partial T2w also successfully rated this plaque as type IV-Va. However, the spatial distribution and percentage area of the lipid core by MRI was quite different from the pathologist’s segmentation and yielded a pixel-by-pixel accuracy of only 45\%. With these 3 MR contrast weightings, MRI underestimated the size of the lipid-rich necrotic core and, in general, placed the necrotic core farther from the lumen.

Discussion

This study provides proof-of-principle that the composition of atherosclerotic plaques determined by high-resolution MRI accurately reflects lesion composition defined by histopathological examination. To the best of our knowledge, this is the first report on an automated classifier trained on 1 ex vivo data set and then successfully validated against an independent data set. A key feature of our automated classification method is that no observer interpretation of the MR images is needed after initial training of the classifier. We believe that validated automated classification algorithms, such as the one presented in this article, could be the first step in achieving maximum reproducibility and reliability in longitudinal in vivo plaque characterization studies, a goal already under investigation by other groups.\textsuperscript{15,16}

In this study, statistical analysis confirmed that the standard 3-contrast technique (PDw, T1w, and partial T2w) does not perform as well as the combination of PDw, T1w, and Dw. This suggests that significant improvement in overall accuracy of plaque characterization may be achieved if practical in vivo diffusion imaging protocols can be developed for high-resolution plaque imaging. Previous studies have identified both T2w and Dw as useful contrasts for identifying a necrotic core.\textsuperscript{5,14,17} However, these studies did not directly compare T2w and Dw; therefore, the relative advantage of Dw over T2w has not been clear until now.

Although the current study was limited to ex vivo analysis, it would not have been feasible to attempt this in an in vivo setting. The use of endarterectomy specimens allowed for the careful study of human atherosclerotic lesions with MR sequences not commonly used in clinical imaging of atherosclerosis (eg, Dw imaging), elucidating the potential value in additional investigation of these sequences. For example, a recent ex vivo study demonstrated optimal contrast between thrombosis and vessel wall with Dw MRI.\textsuperscript{18} In vivo plaque imaging will naturally incur decreased resolution, motion...
artifacts, and greater variability in MR signal characteristics, all of which will complicate the implementation and interpretation of both manual and automated classification methods. Therefore, the use of ex vivo data is a critical first step in the establishment and validation of sophisticated MR pulse sequences and plaque characterization algorithms.

The use of ex vivo specimens also carries with it some fundamental limitations, and extrapolation of our results to the in vivo setting must be approached with caution. Because of the differences in image quality mentioned above, we have not accounted for the limited resolution and signal-to-noise ratio, motion artifact, and other imperfections intrinsic to in vivo MRI. In addition, high-quality $D_w$ images are more difficult to achieve in vivo, especially in the vicinity of the carotid bifurcation where flow characteristics and motion are significant factors.

We previously reported a sensitivity of 83.9% for necrotic core using a minimum-distance-to-means classifier. Despite using the maximum likelihood classifier (a more sophisticated algorithm) in the current study, the sensitivity for necrotic core dropped to 67%. We believe this difference is because of the addition of hemorrhage as a tissue type. Hemorrhage and necrotic core have similar MR signal characteristics and, thus, were difficult to separate with accuracy. Several articles in the field of MR characterization of atherosclerotic plaque have reported values for sensitivity and specificity higher than those described here. However, these studies analyzed atherosclerotic plaques in quadrants, and the simple presence or absence of the plaque components was then assessed with MRI. Such methodology does not evaluate the ability of MR to precisely determine the spatial location and size of plaque components. The pixel-by-pixel method does provide such spatial information and leads to a different and far more stringent definition of MR classification accuracy, with correspondingly lower values for sensitivity, specificity, and overall accuracy.

Acknowledgments

S.E.C. was the recipient of funding from the Canadian Institutes of Health Research. B.K.R. receives salary support from the Barnett-Ivey Heart and Stroke Foundation of Ontario Endowed Chair award, and this work was enabled by operating grants from the Canadian Institutes for Health Research (grants MT-11540 and GR-14973), the Ontario Research and Development Challenge Fund, and General Electric Medical Systems. R.A.H. is supported by a Canada Research Chair (Tier I) in Human Genetics and by a Career Investigator award from the Heart and Stroke Foundation of Ontario and holds operating grants from the Canadian Institutes for Health Research and the Heart and Stroke Foundation of Ontario. V.B. was supported by a Heart and Stroke Foundation of Ontario research fellowship award.

References

Validation of Automatically Classified Magnetic Resonance Images for Carotid Plaque Compositional Analysis
Sharon E. Clarke, Vadim Beletsky, Robert R. Hammond, Robert A. Hegele and Brian K. Rutt

*Stroke*. 2006;37:93-97; originally published online December 8, 2005;
doi: 10.1161/01.STR.0000196985.38701.0c

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/37/1/93

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org//subscriptions/