A Critical Appraisal of the Performance, Reporting, and Interpretation of Studies Comparing Carotid Plaque Imaging With Histology

J.K. Lovett, MRCP; J.N.E. Redgrave, MRCP; P.M. Rothwell, FRCP

Background and Purpose—Carotid plaque instability is an important determinant of stroke risk. There are now a number of different imaging techniques that provide information on carotid plaque morphology. However, it is unclear how they compare with one another or whether they can reliably assess plaque instability. Studies comparing imaging with pathology have shown highly variable results, even for similar imaging techniques. This may be because of variable pathology techniques rather than differences in imaging.

Methods—We performed a systematic review of studies that compared carotid imaging with histology of the excised plaque published between January 1995 and September 2004. We assessed the quality and comparability of these studies. In particular, we determined which histology methods were used and whether observer reproducibility of the histology assessment was reported.

Results—Among 73 eligible studies, histological methods were poorly reported and highly variable; 23% reported reproducibility data for imaging and only 12% reported reproducibility data for histology. Of 29 studies that reported quantitative results of blinded comparisons, there were methodological deficiencies and the results were highly variable. No study considered the extent to which the lack of reproducibility influenced the imaging-pathological correlations reported.

Conclusions—Pathological correlation in studies of carotid plaque imaging cannot be reliably interpreted or compared because of incomparable and poorly reported histology methods. We make recommendations for the performance, reporting, and interpretation of imaging–pathological correlation studies and highlight the need for consensus guidelines. (Stroke. 2005;36:1085-1091.)

Key Words: arteriosclerosis • carotid artery plaque • cerebral angiography • pathology

Carotid artery atherosclerosis is the underlying cause of ≈10% to 20% of strokes.1 Degree of stenosis at the carotid bifurcation is the most widely used radiological predictor of stroke risk on medical treatment and of likely benefit from carotid endarterectomy.2-3 However, pathological studies have shown that plaque instability, rather than the extent of disease, is the main cause of acute thromboembolic events.4-7 Acute myocardial infarction is associated with a large lipid core, a thin ruptured cap, and adherent thrombus.4,5 and similar features have been found in pathological studies of carotid arteries from patients with recent stroke.6,7 Furthermore, 2 large prospective studies have shown that angiographic carotid plaque ulceration is a strong independent predictor of ipsilateral ischemic stroke on follow-up, and of increased benefit from endarterectomy.8,9 However, ultrasound, magnetic resonance imaging, and computed tomography are now widely used to assess carotid artery disease. Although these techniques have the advantage of being less invasive than intra-arterial angiography, it is important that they are validated against angiography and pathological specimens.

There have been a large number of published studies on the imaging–pathological correlation of unstable carotid plaque. The majority has imaged the carotid arteries of patients awaiting endarterectomy and used the histological assessment of the excised plaque as the “gold standard” for comparison. However, the results have been highly variable, even when similar methods of imaging have been used, for example the Gray–Weale scale of ultrasound echogenicity.10-13 One explanation is that variability arises not from the imaging methods, but from the pathological assessments. Histopathological assessments, such as staging of cervical, prostatic, ovarian, and lung neoplasia, commonly show considerable variability, even between experienced pathologists.14-18 In a study of the reproducibility of a histological assessment of 60 carotid plaques, we found significant intra-observer and inter-observer variability even with simple and well-defined assessments, eg, for presence of rupture, the \( \kappa \) values for...
intra-observer and inter-observer variability were 0.68 and 0.43, respectively. Moreover, we demonstrated that there is considerable variation between sections taken from different positions in the same plaque. It is important, therefore, that imaging versus pathology agreement within studies and variability in results between studies of the same imaging modality are interpreted in the light of the reproducibility of the histological assessments and any imprecision introduced by the pathology methods, as well as the properties of the imaging technique.

Therefore, we reviewed the literature on studies that compared carotid imaging with histology. In particular, we aimed to determine which histology methods were used and whether the histology methods were described in sufficient detail to make valid comparisons between studies. We also assessed whether observer reproducibility of the imaging and histology assessments was reported, and we used our own data to demonstrate the influence of observer variability on an imaging–pathological comparison.

Materials and Methods
We searched for studies that compared imaging of carotid plaque morphology with histology of plaque excised at endarterectomy or postmortem, published between January 1995 and September 2004. Our search strategy was as follows: (1) we searched Medline using combinations of the search terms “carotid,” “atheroma,” “atherosclerosis,” “plaque,” “histolog,” “patholog,” “angiograph,” “ultrason,” “Doppler,” “duplex,” “radio,” and “imaging”; (2) searches were restricted to articles in English using human subjects; (3) we hand-searched the contents pages of the 4 journals from which articles were most frequently found in the electronic search; and (4) we examined reference lists from all articles.

For each study we recorded the number of carotid arteries imaged and the modality and method of radiological assessment. We noted the method of pathological specimen preparation frequency, number, and plane of sectioning and staining, the method of histological data collection recorded features and system of grading for each feature, and the method of comparison including whether there was blinding. We also noted whether the authors assessed reproducibility of the imaging and histology assessment or whether there was any reference to previously used histological classifications.

Results
We found 73 eligible studies. Articles were most frequently found in the journals Stroke (n=11), Circulation (n=10), European Journal of Vascular and Endovascular Surgery (n=6), and Journal of Vascular Surgery (n=4). All studies used histology of excised carotid plaque, from endarterectomy in 65 studies, from cadavers in 7 studies, and from both sources in 1 study as the “gold standard.” Nineteen (26%) studies used carotid plaques from patients with a history of ipsilateral ischemic symptoms, 26 (36%) used plaques from both symptomatic and asymptomatic patients, but in 28 (38%) studies the symptomatic status of the patients was not reported. The number of carotid plaques in the studies ranged from 2 to 270 (median, 24; interquartile range, 15 to 52).

Imaging
Thirty-nine studies used carotid ultrasound as the imaging modality, 4 used intra-arterial angiography, and 30 used magnetic resonance imaging. There were 3 studies that did not provide any imaging method. The other entries were a combination of transcranial Doppler, transcranial Doppler and magnetic resonance angiography, transcranial Doppler and magnetic resonance imaging, transcranial Doppler and computed tomography, and computed tomography and magnetic resonance imaging. In 28 (38%) studies the symptomatic status of the patients was not reported.

Histological Methods
There was great variation in the preparation of the plaque for histological assessment. Forty-eight (66%) studies used transverse sections. The frequency of sectioning ranged from 0.2 mm to 5 mm (median, 3 mm), but was not stated in 52 (71%) studies. Six of the 9 studies assessed reproducibility of the imaging technique. Forty-three (59%) of these studies did not report the time interval between imaging and surgery.

Materials and Methods
We searched for studies that compared imaging of carotid plaque morphology with histology of plaque excised at endarterectomy or postmortem, published between January 1995 and September 2004. Our search strategy was as follows: (1) we searched Medline using combinations of the search terms “carotid,” “atheroma,” “athero-
<table>
<thead>
<tr>
<th>Reference</th>
<th>Plaques, No.</th>
<th>Histology Assessment</th>
<th>Imaging Assessment</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnold 1999&lt;sup&gt;38&lt;/sup&gt;</td>
<td>60</td>
<td>Lipid/fibrous</td>
<td>US echogenicity: 2 or 4 grades</td>
<td>( \kappa = 0.12 \times 4 \times 4 )</td>
</tr>
<tr>
<td>Schulte 2000&lt;sup&gt;42&lt;/sup&gt;</td>
<td>46</td>
<td>Lipid/fibrous, hemorrhage, calcification, thrombus</td>
<td>US echogenicity: 6 grades</td>
<td>Range of sensitivities and specificities for each grade: 35%–97%</td>
</tr>
<tr>
<td>Arnold 2001&lt;sup&gt;54&lt;/sup&gt;</td>
<td>50</td>
<td>Lipid/fibrous, hemorrhage, calcification</td>
<td>US echogenicity: 2 grades</td>
<td>( \kappa = 0.09 \times 2 \times 2 )</td>
</tr>
<tr>
<td>Gronholdt 2001&lt;sup&gt;64&lt;/sup&gt;</td>
<td>38</td>
<td>Lipid/fibrous, hemorrhage, calcification, thrombus</td>
<td>US echogenicity: continuous scale; CT: Hounsfield units</td>
<td>CT correlated with calcium, no significant correlations between US and histology</td>
</tr>
<tr>
<td>Droste 1997&lt;sup&gt;98&lt;/sup&gt;</td>
<td>29</td>
<td>Lipid/fibrous, hemorrhage, thrombus</td>
<td>US echogenicity: 2 or 4 grades</td>
<td>No significant relationships</td>
</tr>
<tr>
<td>Lammie 2000&lt;sup&gt;81&lt;/sup&gt;</td>
<td>22</td>
<td>Lipid/fibrous, hemorrhage, cap thickness, rupture, thrombus</td>
<td>US echogenicity: 4 grades</td>
<td>( \kappa = 0.53 \times \text{cap thickness}, \text{other comparisons} )</td>
</tr>
<tr>
<td>Kardoulas 1996&lt;sup&gt;67&lt;/sup&gt;</td>
<td>36</td>
<td>Lipid/fibrous, hemorrhage, calcification, rupture</td>
<td>US echogenicity: 4 grades</td>
<td>Significant association only with fibrous tissue, ( P = 0.04 )</td>
</tr>
<tr>
<td>Gaunt 1996</td>
<td>50</td>
<td>Lipid/fibrous, rupture, thrombus</td>
<td>US echogenicity: 4 grade scale and surface morphology</td>
<td>66% agreement with ulceration and thrombus</td>
</tr>
<tr>
<td>Sitzer 1996&lt;sup&gt;94&lt;/sup&gt;</td>
<td>37</td>
<td>Rupture</td>
<td>US: Surface morphology</td>
<td>Sensitivity 33%, specificity 76%</td>
</tr>
<tr>
<td>Denzel 2003&lt;sup&gt;63&lt;/sup&gt;</td>
<td>107</td>
<td>Lipid/fibrous</td>
<td>US echogenicity 3 grades</td>
<td>Overall agreement 46.7%, ( \kappa = 0.168, P = 0.013 )</td>
</tr>
<tr>
<td>Denzel 2004&lt;sup&gt;45&lt;/sup&gt;</td>
<td>92</td>
<td>Hemorrhage, lipid/fibrous, calcification</td>
<td>US echogenicity and calcium score on CT/XR</td>
<td>GSM ( \kappa = -0.088, P = 0.22 ); CT calcium ( P = 0.001 ); XR calcium ( \kappa = 0.475, P = 0.001 )</td>
</tr>
<tr>
<td>Tegos 2000&lt;sup&gt;39&lt;/sup&gt;</td>
<td>71</td>
<td>Hemorrhage, lipid/fibrous, thrombus</td>
<td>US echogenicity 2 grades</td>
<td>No correlation with lipid core size ( P = 0.37 ), ( r = 0.1 ) or calcification ( P = 0.65 ), low GSM associated with hemorrhage ( P = 0.04 )</td>
</tr>
<tr>
<td>Rakebrandt 2000&lt;sup&gt;28&lt;/sup&gt;</td>
<td>10</td>
<td>Hemorrhage, lipid, elastin, fibrin, calcification</td>
<td>US trained texture classifier: 5 classes</td>
<td>US overestimated calcium content by 45% and lipid by 25%; hemorrhage underestimated by 63%</td>
</tr>
<tr>
<td>Hatsukami 2000&lt;sup&gt;32&lt;/sup&gt;</td>
<td>36</td>
<td>Hemorrhage, rupture, thrombus, cap thickness</td>
<td>MRI: cap thickness and rupture</td>
<td>Rupture ( \kappa = 0.85 ), thick vs thin cap ( \kappa = 0.83 )</td>
</tr>
<tr>
<td>Yuan 2001&lt;sup&gt;33&lt;/sup&gt;</td>
<td>18</td>
<td>Lipid/fibrous, hemorrhage, AHA grade</td>
<td>MRI features</td>
<td>Sensitivity 85%, specificity 92%, ( \kappa = 0.69 )</td>
</tr>
<tr>
<td>Cai 2002&lt;sup&gt;20&lt;/sup&gt;</td>
<td>52</td>
<td>AHA grade</td>
<td>MRI: AHA grade equivalents</td>
<td>Agreement 80%</td>
</tr>
<tr>
<td>Serfati 2001&lt;sup&gt;17&lt;/sup&gt;</td>
<td>18</td>
<td>Lipid/fibrous, calcification, AHA grade</td>
<td>MRI: AHA grade equivalents</td>
<td>MRI identified fibrocalcific plaques but otherwise poor results</td>
</tr>
<tr>
<td>Shinnar 1999&lt;sup&gt;40&lt;/sup&gt;</td>
<td>22</td>
<td>Lipid/fibrous, thrombus, calcification, AHA grade</td>
<td>MRI features</td>
<td>Sensitivities 84–100%, specificities 95–100%</td>
</tr>
<tr>
<td>Clarke 2003&lt;sup&gt;41&lt;/sup&gt;</td>
<td>8</td>
<td>Lipid/fibrous, calcification, loose connective tissue</td>
<td>MRI: pixel analysis of plaque appearances; micro-CT for colocalization of calcium</td>
<td>Sensitivity 80.4–97.8%, specificity 75–98.9% for detecting plaque constituents</td>
</tr>
<tr>
<td>Mitsumori 2003&lt;sup&gt;42&lt;/sup&gt;</td>
<td>18</td>
<td>Unstable cap, rupture, cap thickness</td>
<td>MRI signal intensity</td>
<td>Sensitivity 0.81, specificity 0.90</td>
</tr>
<tr>
<td>Kerwin 2003&lt;sup&gt;47&lt;/sup&gt;</td>
<td>16</td>
<td>Microvessels—fraction of plaque area</td>
<td>MRI dynamic contrast enhancer fractional blood volume</td>
<td>Correlation coefficient ( r = 0.80, P = 0.001 )</td>
</tr>
<tr>
<td>Trivedi 2004&lt;sup&gt;48&lt;/sup&gt;</td>
<td>23</td>
<td>Hemorrhage, lipid/fibrous, calcification</td>
<td>MRI signal intensity and area of plaque constituents</td>
<td>Mean differences ( t = 2 ) SD lipid core content ( P = 0.076 ), ( r = 0.75 )</td>
</tr>
<tr>
<td>Chu 2004&lt;sup&gt;49&lt;/sup&gt;</td>
<td>24</td>
<td>Hemorrhage age—fresh, recent, old</td>
<td>MRI signal intensity; hemorrhage age: 3 grades</td>
<td>Sensitivity 90%, specificity 74% for hemorrhage detection, ( \kappa = 0.44 ) and 0.86 (2 observers for hemorrhage age)</td>
</tr>
<tr>
<td>Moody 2003&lt;sup&gt;40&lt;/sup&gt;</td>
<td>63</td>
<td>AHA grade 6</td>
<td>MRI signal intensity on direct thrombus imaging</td>
<td>Sensitivity 84%, specificity 84%, negative predictive value 70%</td>
</tr>
<tr>
<td>Ternery 2003</td>
<td>26</td>
<td>Microphages, smooth muscle cells, cap thickness</td>
<td>OCT signal intensity</td>
<td>OCT density ( r = 0.84, P = 0.001 ); macrophage density ( t = 0.56, P = 0.005 )</td>
</tr>
<tr>
<td>Yabushita 2002&lt;sup&gt;28&lt;/sup&gt;</td>
<td>105</td>
<td>Lipid/fibrous, calcification</td>
<td>OCT</td>
<td>Agreement of 52%–91%</td>
</tr>
<tr>
<td>Oliver 1999&lt;sup&gt;76&lt;/sup&gt;</td>
<td>9</td>
<td>Lipid/fibrous, hemorrhage, rupture, calcification, cap thickness</td>
<td>CT: calcification, surface morphology</td>
<td>Sensitivity 60%, specificity 74%</td>
</tr>
<tr>
<td>Walker 2002&lt;sup&gt;27&lt;/sup&gt;</td>
<td>55</td>
<td>Lipid/fibrous, hemorrhage, rupture</td>
<td>CT: surface morphology</td>
<td>detected 8/9 thrombi, but no association with lipid pools</td>
</tr>
<tr>
<td>Manca 2001&lt;sup&gt;39&lt;/sup&gt;</td>
<td>22</td>
<td>Lipid/fibrous, hemorrhage, rupture, thrombus, calcification</td>
<td>Platelet scintigraphy: thrombus and lipid</td>
<td></td>
</tr>
</tbody>
</table>

CT indicates computed tomography; GSM, grayscale median; OCT, optical coherence tomography; US, ultrasound; XR, X-ray.

**Results of the Imaging–Pathological Comparisons**

Of the 73 studies, 29 (40%) reported blinded comparisons between imaging and histology and gave quantitative results on diagnostic accuracy, eg, sensitivity, specificity, or \( \kappa \) values (Table 1). The overall results were highly variable even when the same imaging modality was studied. For example, for ultrasound appearance of plaque, reported sensitivity and specificity for histological plaque appearance ranged from 33% to 97%, and reported \( \kappa \) values ranged from 0.08 to 0.53 (Table 1). Findings of studies using magnetic resonance imaging were more consistent, showing overall stronger correlations with histology (Table 1). However, it was not possible to pool results of the studies, because of incompa-
table and poorly reported methods. One small study\textsuperscript{20} (n=10) assessed the effect of storage temperature and plaque decalci-
ification on the results of imaging–pathologic comparisons, but no study considered the effect of poor reproducibility of
histological assessments.

Discussion
This review found 73 studies, the majority from journals with a high “impact factor,” that compared carotid plaque imaging
with carotid plaque histology. However, the results were highly variable, even between studies that used the same
imaging criteria, such as the Gray–Weale scale of ultrasound
echogenicity,\textsuperscript{10–13} which has been shown to be reproducible.\textsuperscript{91}
Direct comparisons between studies were not possible because of incomparable methods, which were often not shown
to be reproducible.

We have recently published data on angiographic plaque
surface morphology and histological appearance of plaque
instability from 128 carotid arteries and demonstrated the
influence of intra-observer reproducibility on the results of an
imaging–pathological comparison. The methods have been
described in detail previously\textsuperscript{10,92,93} but briefly, the intra-
observer and inter-observer reproducibility for angiographic
appearance of irregular or ulcerated plaque versus smooth
plaque was assessed on 50 and 1000 angiograms, respect-
atively. Two independent pathologists calculated intra-
observer and inter-observer reproducibility for a well-defined
simple histology assessment of plaque instability. Blinded
comparisons were made between histology and angiographic
plaque surface morphology.

The agreement between presence of irregularity or ulcer-
ation on angiogram with unstable plaque on histology was
significant (P=0.01), but the calculated $\kappa$ was only 0.22
(95% CI, 0.05 to 0.39). The intra-observer reproducibility $\kappa$
values for angiographic assessments were 0.56 (95% CI, 0.2
to 0.9), 0.60 (95% CI, 0.2 to 0.9), and 0.67 (95% CI, 0.3 to
1.0) for 3 observers, and the inter-observer $\kappa$ value was 0.56
(95% CI, 0.53 to 0.59).\textsuperscript{93} For histological assessments, the
intra-observer $\kappa$ values ranged from 0.35 (95% CI, 0 to 0.7)
for presence of rupture to 0.89 (95% CI, 0.8 to 1) for presence
of thrombus.\textsuperscript{19} The intra-reproducibility and inter-
reproducibility $\kappa$ values for the composite histological assess-
ment of “unstable” plaque were 0.69 (95% CI, 0.5 to 0.9) and
0.40 (95% CI, 0.2 to 0.6), respectively.

The calculation of an actual expected maximum $\kappa$ value for
angiography–histology comparisons based on observer re-
producibility data are highly complex, but can be no greater
than the reproducibility of each individual assessment. There-
fore, we recommend that studies of this type report intra-
observer and inter-observer reproducibility for the histologi-
cal assessments as well as those for the imaging technique,
and the results of the imaging–pathological comparison
should be interpreted in the light of these results.

It has consistently been shown in population-based studies
that the risk of recurrent stroke after initial ischemic symp-
toms declines with time.\textsuperscript{94,95} Moreover, the early risk is
highest in patients with large artery atherosclerosis compared
with the other etiological subtypes.\textsuperscript{96} One explanation for this
is that unstable plaques heal with time. We have shown
previously that plaques from patients with ischemic symp-
toms within 30 days of endarterectomy are more unstable on
histology than from those with less recent symptoms.\textsuperscript{97} Thus,
in studies comparing in vivo appearances of carotid plaque
with histology of the excised specimens, it is essential to state
whether the plaques have been symptomatic. Furthermore,
the time interval between imaging and histology should
ideally be minimized and kept consistent within studies.
Twenty-eight (39%) of the studies in this review did not
comment on the symptomatic status of their patients; thus, the
application of their findings to clinical practice is severely
limited.

Several studies in this review reported only 1 or 2 histo-
logical features despite the existence of validated grading
scales for the severity of atherosclerosis.\textsuperscript{90,98} Although in
some cases this may have been deliberate, reflecting the
limitations of the imaging modality concerned to detect
certain characteristics, it raises the possibility of selective
reporting of histological features. Therefore, we recommend
that unless the reasons for reporting selective characteristics
are clearly justified, the following well-recognized features
of unstable plaque should be described: hemorrhage,
rapture, thrombus, lipid core-to-fibrous tissue ratio, and minimum cap
thickness; this is to avoid the potential for selective publication

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
\textbf{TABLE 2. Recommendations for the Performance and Reporting of Studies of Carotid Plaque Imaging vs Histology} \\
\hline
Sample size calculations should be performed and should take into account the reproducibility of the assessments studied. \\
$\kappa$ values for the intra-observer and inter-observer reproducibility of imaging and histological assessments should be quoted and the results interpreted in the light of these \\
The numbers of symptomatic and asymptomatic plaques should be reported as well as the time since last symptoms \\
The time interval between in vivo plaque imaging and plaque excision should be reported and kept to a minimum; this is particularly important in the study of recently symptomatic plaques \\
Sufficient detail of the plaque processing methods, particularly whether plaques were decalcified before sectioning should be given \\
Histological assessments should be made blind to imaging and clinical characteristics—when necessary using a second observer to colocalize imaging slices with histology \\
There should be detailed reporting of histological methods including position, plane, number, and frequency of sections; as a minimum, we recommend that sections are taken from the bifurcation or point of maximal stenosis, and from 3 mm either side \\
When there is no clear reason to report a specific histological feature, the following features of unstable plaque should be described: hemorrhage, rupture, thrombus, lipid core-to-fibrous tissue ratio, and minimum cap thickness; this is to avoid the potential for selective publication \\
\hline
\end{tabular}
\end{table}
form the histological assessments and the colocalization, each blind to imaging, and to the results of the other observer.

The position, plane, number, and frequency of histology sections taken from each carotid plaque were infrequently reported in this review. We have previously shown that there is considerable variability in histological features between transverse sections taken at 3-mm intervals along the length of a plaque, but that 3 sections taken from the bifurcation point and from 3 mm proximal and 3 mm distal to this point are sufficient to detect almost all features found on a larger number of sections from the same plaque.19 We therefore recommend that this method be adopted as a minimum, so that important features seen on imaging are not missed by the histology assessment.

The median number of plaques studied in this review was 24 (interquartile range, 15 to 52). In a hypothetical study of this size, with a 33% prevalence of the investigated characteristic and 75% imaging–pathological agreement, the 95% CI of the \( \kappa \) statistic would be 0.1 to 0.8 (ranging from poor to very good agreement). Therefore, larger sample sizes are required if the degree of imaging–pathological associations are to be demonstrated reliably and differences between imaging techniques are to be detected.

There are several stages in the preparation of carotid plaques for histological analysis. The process of plaque decalcification is frequently performed to remove large calcified nodules before sectioning. We found that few studies reported data on plaque decalcification or other processing methods, even when this could have influenced subsequent imaging–pathological correlations.

Given the shortcomings of published studies and the consequent difficulty in the reliable comparison and interpretation of the results, consensus guidelines on the performance and reporting of these studies are required. We suggest that our recommendations form the basis for these guidelines (Table 2).

Acknowledgments
We acknowledge the Guaran tors of Brain, the Wellcome Trust, and the Medical Research Council for their financial support.

References


60. Wahi K, Masuyama T, Mori H, Maela T, Kinade K, Moriyasu K, Tsujimoto M, Fujimoto K, Koshima N, Matsuura N. Ultrasonic tissue